



# **Universidad Nacional Mayor de San Marcos**

**Universidad del Perú. Decana de América**

**Facultad de Ciencias Biológicas**

**Escuela Profesional de Genética y Biotecnología**

## **Estudio de asociación entre los SNPs rs4073 y rs2227543 del gen interleucina 8 (IL-8) y presencia de paraparesia espástica tropical (PET) en individuos peruanos portadores del virus HTLV-1**

### **TESIS**

**Para optar el Título Profesional de Biólogo Genetista  
Biotecnólogo**

### **AUTOR**

**Jorge RUA FERNANDEZ**

### **ASESORES**

**Monica Yolanda PAREDES ANAYA**

**Wilser Andrés GARCÍA QUISPES**

**Michael John TALLEDO ALBÚJAR**

**Lima, Perú**

**2017**



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

## Referencia bibliográfica

---

Rua, J. (2017). *Estudio de asociación entre los SNPs rs4073 y rs2227543 del gen interleucina 8 (IL-8) y presencia de paraparesia espástica tropical (PET) en individuos peruanos portadores del virus HTLV-1*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Biológicas, Escuela Profesional de Genética y Biotecnología]. Repositorio institucional Cybertesis UNMSM.

---



**UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS**  
(Universidad del Perú, DECANA DE AMÉRICA)

**FACULTAD DE CIENCIAS BIOLÓGICAS**

**ACTA DE SESIÓN PARA OPTAR AL TÍTULO PROFESIONAL DE  
BIÓLOGO GENETISTA BIOTECNÓLOGO  
(MODALIDAD: SUSTENTACIÓN DE TESIS)**

Siendo las 14:35 horas del 08 de agosto de 2017, en el Salón de Grados de la Facultad de Ciencia Biológicas y en presencia del jurado formado por los profesores que suscriben, se dio inicio a la sesión para optar al Título Profesional de Biólogo Genetista Biotecnólogo de **JORGE RUA FERNANDEZ**.

Luego de dar lectura y conformidad al expediente N° 005-EPGB-2017, el titulando expuso su tesis: **"ESTUDIO DE ASOCIACIÓN ENTRE LOS SNPs rs4073 Y rs2227543 DEL GEN INTERLEUCINA 8 (IL-8) Y PRESENCIA DE PARAPARESIA ESPÁSTICA TROPICAL (PET) EN INDIVIDUOS PERUANOS PORTADORES DEL VIRUS HTLV-1"** y el Jurado efectuó las preguntas del caso calificando la exposición con la nota 18...., calificativo: aprobado con mención honorosa.

Finalmente, el expediente será enviado a la Escuela Profesional de Genética y Biotecnología, y al Consejo de Facultad para que se apruebe otorgar el **Título Profesional de Biólogo Genetista Biotecnólogo** a **JORGE RUA FERNANDEZ** y se eleve lo actuado al Rectorado para conferir el respectivo título, conforme a ley.

Siendo las 15:50 horas se levantó la sesión.

Ciudad Universitaria, 08 de agosto de 2017.

Dra. EGMA MAYTA HUATUCO  
(PRESIDENTA)

Dra. MONICA PAREDES ANAYA  
(ASESORA)

Blga. MARGARITA VELASQUEZ REINOSO  
(MIEMBRO)

Blgo. ERASMO COLONA VALLEJOS  
(MIEMBRO)

## DEDICATORIA

A mis padres, Flor María Fernández Ramírez y Olimpio Rúa Linares por siempre darme su apoyo incondicional, su amor y su confianza para dejarme volar cada vez más alto. A mis hermanos, Omar Rúa Fernández y Oliver Rúa Fernández, que son mi ejemplo a seguir como personas y profesionales. Gracias, no solo por ayudarme a aprender a caminar por segunda vez, sino por darme un hogar lleno de cariño, risas y dedicación. Me encanta poder dar los mejores pasos de mi vida junto a ustedes.

A mis sobrinos Piero y Camila, que son una luz y una imagen lo bueno que hay en el mundo y por lo que también quiero hacer con mi carrera un lugar mejor para ellos.

A mi abuelita Guillermina por su inmenso cariño y a mis abuelitos José Irene, Marcelina y Fortunato que siempre están en mi mente y en mi corazón.

## AGRADECIMIENTOS

A mi asesora Mónica Yolanda Paredes Anaya, que nunca dejó de confiar en mí incluso en los momentos más álgidos de este proceso. Gracias por darse tiempo de donde no había y mantener su compromiso en ayudarme a culminar esta etapa tan importante en mi vida profesional. Su optimismo, su sinceridad y las ganas de seguir ayudándome a alcanzar cada nueva meta que me proponía, hacen que me invada una gratitud eterna hacia usted.

A mi coasesor Wilser Andrés García Quispes, por introducirme al mundo de la estadística y la programación dirigiendo mi norte siempre a encontrar el aspecto biológico importante de todo lo que investigaba, porque como suele decirme “el resto son solo números”. Debo mencionar que su buen sentido del humor es tan peculiar como la coincidencia en nuestros temas de investigación, lo que me lleva a pensar que conocerlo no fue algo tan aleatorio que digamos, sino más bien que fue y sigue siendo muy significativo para mí, a pesar de que reniegue de ese término. Mi eterno agradecimiento para usted profesor.

A mi coasesor externo Michael John Talledo Albújar, a quien considero un segundo padre, me abrió las puertas de su laboratorio y a un entrenamiento que me serviría para alcanzar hasta el momento los más grandes logros en mi vida. “Uno mismo se pone sus propios límites” y gracias a usted he aprendido a superar varios que yo mismo creía imposibles de lograr. Es una gran satisfacción encontrar en un excelente profesional y extraordinaria persona, también a un gran amigo.

A la profesora Olga Hilda Bracamonte Guevara y al grupo de investigación ADCA con los que inicié en su laboratorio esta aventura de investigar fuera de las aulas y enfocar el estudio de la genética de un modo más experimental.

A mi familia del laboratorio de HTLV-1 que solo traen sonrisas en mis recuerdos. Gracias por hacer del ambiente de trabajo el lugar perfecto para aprender, discutir artículos, reírnos y apreciar la vida por el hecho de habernos conocido. Francesca, Giovanni, Jason, Andrés, Jessica, Romina, Maryam, Jean Pierre, Dilan, Gianina, Gabriela y Astrid tienen un lugar eterno en mi corazón.

A mi base de genética y Biotecnología “Alexandra Elbakyan”, más conocida como “Base 11”, por ser no solo un ambiente competitivo y de constante dedicación al estudio, sino también por la amistad y la fraternidad que hasta ahora nos mantiene unidos demostrando lo mejor de cada uno en esta hermosa carrera que escogimos.

A todos ustedes, gracias.

## ABREVIATURAS

ADN	Ácido desoxirribonucleico
ADNc	ADN complementario
AIC	Criterio de información de Akaike
AIM	Marcador informativo de ancestría
ARN	Ácido ribonucleico
ARNm	ARN mensajero
CPV	Carga proviral
DE	Desviación estándar
dNTPs	Desoxirribonucleótidos trifosfato
<i>Forward</i>	Sentido 5' a 3'
HAM/TSP	HTLV-1-associated myelopathy/tropical spastic paraparesis
H <sub>2</sub> O	Agua
IkB $\alpha$	Inhibidor del factor nuclear del potenciador del gen del polipéptido ligero kappa en células B
IKK- $\alpha$	Inhibidor de factor nuclear quinasa kappa-B subunidad alfa
IKK- $\beta$	Inhibidor de factor nuclear quinasa kappa-B subunidad beta
IKK- $\gamma$	Inhibidor de factor nuclear quinasa kappa-B subunidad gama
IL-8	Interleucina 8
LLTA	Leucemia de linfocitos T adultos
LTR	Repetición terminal larga
MAF	Frecuencia del alelo menos común
MgCl <sub>2</sub>	Cloruro de magnesio
mM	Milimolar
NF-kB	Factor nuclear potenciador de las cadenas ligeras kappa de las células B activadas.
ng	Nanogramo



pb	Pares de bases
OR	<i>odds ratio</i>
PCR	Reacción en cadena de la polimerasa
PET	Paraparesia espástica tropical
RelA	Factor de transcripción p65
<i>Reverse</i>	Sentido 3' a 5'
Rx	Reacción
SNP	Polimorfismo de nucleótido simple
μl	Microlitro
μM	Micromolar

## ÍNDICE GENERAL

1. INTRODUCCIÓN.....	1
2. MARCO TEÓRICO.....	3
2.1 Virus HTLV.....	3
2.2 Epidemiología .....	3
2.3 Estructura genómica del virus HTLV-1.....	4
2.4 Infección y transmisión de HTLV-1 .....	6
2.5 Enfermedades asociadas a HTLV-1.....	6
2.6 Paraparesia espástica tropical (PET).....	7
2.7 Vía NF- $\kappa$ B en PET .....	7
2.8 Interleucina 8 y la vía NF- $\kappa$ B.....	9
2.9 Factores de riesgo conocidos para el desarrollo de PET .....	11
2.9.1 Carga proviral.....	11
2.9.2 Edad .....	12
2.9.3 Sexo.....	12
2.9.4 La genética del hospedero.....	13
2.10 Análisis de asociación genética .....	14
2.10.1 Tipos de estudios de asociación genética .....	14
2.10.2 Marcadores de ancestría.....	15
2.11 SNPs en los estudio de asociación .....	16
2.12 Antecedentes de estudios de asociación de SNPs en <i>IL-8</i> .....	16
2.13 La paraparesia espástica tropical en el Perú.....	18
2.14 Justificación del estudio.....	19
3. HIPÓTESIS Y OBJETIVOS .....	21
3.1 Hipótesis .....	21
3.2 Objetivos.....	21
3.2.1 General. ....	21
3.2.2 Específicos .....	21
4. MATERIALES Y MÉTODOS.....	22
4.1 Declaración ética.....	22
4.2 Obtención de la muestra .....	22

4.2.1	Criterios de inclusión.....	22
4.2.2	Criterios de exclusión.....	23
4.3	Extracción de ADN .....	23
4.4	Base de datos .....	23
4.5	Selección de SNPs .....	24
4.6	Genotipificación .....	25
4.6.1	Diseño de cebadores.....	26
4.6.2	PCR .....	28
4.6.3	Electroforesis.....	29
4.7	Análisis estadístico.....	29
4.7.1	Análisis descriptivo de variables categóricas y numéricas.....	30
4.7.2	Correlación de variables numéricas .....	30
4.7.3	Evaluación de SNPs .....	30
4.7.4	Análisis de asociación de genotipos.....	31
4.7.5	Análisis de asociación de haplotipos.....	33
4.7.6	Regresión lineal de la carga proviral .....	33
5.	RESULTADOS .....	34
5.1	Variables categóricas.....	34
5.2	Variables numéricas.....	36
5.3	Genotipificación .....	38
5.4	Distribución de CPV por genotipos.....	39
5.4.1	SNP rs2227543.....	39
5.4.2	SNP rs4073 .....	40
5.5	Análisis de asociación .....	41
5.5.1	SNP rs2227543.....	41
5.5.2	SNP rs4073 .....	45
5.6	Análisis de asociación de haplotipos.....	49
5.6.1	Análisis crudo.....	49
5.6.2	Análisis ajustado 1 .....	50
5.6.3	Análisis ajustado 2 .....	50
5.7	Regresión lineal con la variable CPV como respuesta .....	51
6.	DISCUSIÓN .....	52

6.1	<i>Frecuencias alélicas y genotípicas</i> .....	52
6.2	<i>Evaluación de los factores de riesgo</i> .....	53
6.3	<i>Análisis de asociación de genotipos</i> .....	54
6.4	<i>Análisis de asociación de haplotipos</i> .....	58
6.5	<i>Relación entre la carga proviral y las variables evaluadas</i> .....	58
6.6	<i>Consideraciones del estudio de asociación genética</i> .....	60
7.	CONCLUSIONES.....	62
8.	RECOMENDACIONES .....	63
9.	BIBLIOGRAFÍA.....	64
10.	ANEXOS.....	73
10.1	ANEXO 1: Descripción de componente principal eig1 .....	74
10.2	ANEXO 2: Descripción de componente principal eig2 .....	75
10.3	ANEXO 3: Descripción del componente principal eig3 .....	76
10.4	ANEXO 4: Correlación de variables numéricas en el grupo PA .....	77
10.5	ANEXO 5: Correlación de variables numéricas en el grupo PET .....	78
10.6	ANEXO 6: Frecuencias alélicas y genotípicas del SNP rs2227543.....	79
10.7	ANEXO 7: Comparación de las frecuencias genotípicas del grupo PA con las frecuencias reportadas por el proyecto Ensembl del SNP rs2227543 .....	80
10.8	ANEXO 8: Frecuencias alélicas y genotípicas del SNP rs4073 .....	81
10.9	ANEXO 9: Comparación de frecuencias genotípicas del SNP rs4073 con frecuencias reportadas por el proyecto Ensembl.....	82
10.10	ANEXO 10: Genotipificación del SNPs rs2227543 y rs4073.....	83

## ÍNDICE DE TABLAS

Tabla 1. Secuencias de los cebadores del sistema tetraprimers ARMS-PCR .....	27
Tabla 2. Master Mix para PCR de cada SNP .....	28
Tabla 3. Protocolo de amplificación de SNPs .....	29
Tabla 4. Ejemplo de cálculo de OR en un análisis crudo de asociación en el modelo dominante. ....	32
Tabla 5. Efecto del genotipo sobre la enfermedad según sus valores de OR. ....	33
Tabla 6. Distribución de las variables categóricas en los grupos PA y PET. ....	34
Tabla 7. Descripción de las variables numéricas en los grupos PA y PET .....	36
Tabla 8. Análisis crudo de asociación del SNP rs2227543.....	41
Tabla 9. Análisis ajustado 1 del SNP rs2227543.....	42
Tabla 10. Análisis ajustado 2 del SNP rs2227543.....	43
Tabla 11. Análisis crudo de asociación del SNP rs4073.....	45
Tabla 12. Análisis ajustado 1 del SNP rs4073.....	46
Tabla 13. Análisis ajustado 2 del SNP rs4073.....	47
Tabla 14. Análisis crudo de haplotipos.....	49
Tabla 15. Análisis ajustado 1 de haplotipos .....	50
Tabla 16. Análisis ajustado 2 de haplotipos .....	50
Tabla 17. Análisis de regresión lineal.....	51
Tabla 18. Descripción de la variable eig1 .....	74
Tabla 19. Descripción de la variable eig2 .....	75
Tabla 20. Descripción de la variable eig3 .....	76
Tabla 21. Frecuencias alélicas y genotípicas del SNP rs2227543 .....	79
Tabla 22. Frecuencias alélicas y genotípicas del SNP rs4073 .....	81

## ÍNDICE DE FIGURAS

Figura 1. Distribución de las regiones endémicas del virus HTLV-1 a nivel mundial (Gessain y Cassar 2012).....	4
Figura 2. Estructura genómica del virus HTLV-1 (Matsuoka y Jeang 2007).....	5
Figura 3. Activación de la vía canónica y no canónica de NF- $\kappa$ B y activación de la vía Akt mediada por Tax (Matsuoka y Jeang 2007). ....	8
Figura 4. Sitio de unión a NF- $\kappa$ B en la región promotora del gen <i>IL-8</i> (Mori <i>et al.</i> 1998)... ..	10
Figura 5. Expresión del gen <i>IL-8</i> en portadores asintomáticos (AC) y portadores con PET (HAM/TSP) (Talledo 2013). ....	11
Figura 6. Localización de SNPs reportados en el gen <i>IL-8</i> en humanos (Ensembl).....	17
Figura 7. Desequilibrio de ligamiento generado con la herramienta web Ensembl (Yates <i>et al.</i> 2016).....	24
Figura 8. Representación esquemática del método de genotipificación tetraprimer ARMS-PCR (Ye <i>et al.</i> 2001).....	26
Figura 9. Frecuencia de individuos por sexo en cada grupo.....	35
Figura 10. Frecuencia de individuos por Grupo de edad en cada grupo.....	35
Figura 11. Distribución de la variable CPV en cada grupo. ....	37
Figura 12. Distribución de la variable Edad en cada grupo.....	37
Figura 13. Genotipificación del SNP rs2227543 por método tetraprimer ARMS-PCR. ....	38
Figura 14. Genotipificación del SNP rs4073 por método tetraprimer ARMS-PCR.....	38
Figura 15. Distribución de la variable CPV entre los genotipos del SNP rs2227543 en los grupos PA y PET .....	39
Figura 16. Distribución de la variable CPV entre los genotipos del SNP rs4073 en los grupos PA y PET .....	40
Figura 17. Valores de OR e intervalos de confianza en cada uno de los modelos .....	44
Figura 18. Valores de OR e intervalos de confianza en cada uno de los modelos .....	48
Figura 19. Distribución de la variable eig1 entre los grupos PA y PET. ....	74
Figura 20. Distribución de la variable eig2 entre los grupos PA y PET. ....	75
Figura 21. Distribución de la variable eig3 entre los grupos PA y PET. ....	76
Figura 22. Diagrama de correlación entre las variables numéricas en el grupo PA.....	77
Figura 23. Diagrama de correlación entre las variables numéricas en el grupo PET.....	78
Figura 24. Distribución de las frecuencias genotípicas entre los grupos PA y PET del SNP rs2227543. ....	79
Figura 25. Comparación de las frecuencias genotípicas del grupo PA con las del proyecto Ensembl para el SNP rs2227543. ....	80
Figura 26. Distribución de las frecuencias genotípicas entre los grupos PA y PET del SNP rs4073.....	81
Figura 27. Comparación de las frecuencias genotípicas del grupo PA con las del proyecto Ensembl para el SNP rs4073. ....	82
Figura 28. Genotipificación de los SNPs rs2227543 y rs4073 .....	83

## RESUMEN

El virus linfotrópico de células T humanas tipo 1 (HTLV-1) es un retrovirus que genera una respuesta inflamatoria que desencadena una enfermedad neurológica denominada Paraparesia Espástica Tropical (PET). Este proceso inflamatorio se debe a la acción de distintas interleucinas involucradas en detener la predominante infiltración de linfocitos T CD4<sup>+</sup> infectados. La interleucina IL8 (IL-8) es una citocina pro inflamatoria que induce la activación y migración de neutrófilos hacia los sitios de infección. Esta proteína es continuamente expresada en células infectadas con HTLV-1. El objetivo de este estudio fue evaluar la asociación estadística en un estudio caso-control entre dos SNPs rs2227543 y rs4073 del gen *IL-8*, con la presencia de PET en individuos portadores del virus HTLV-1. No se observan diferencias estadísticamente significativas en la distribución de las frecuencias genotípicas y haplotípicas entre los grupos control y PET. Sin embargo, el genotipo T/T de rs2227543 muestra una tendencia a un efecto de riesgo (OR=3.40 [0.98 a 11.84], *p-value*=0.048), mientras que el genotipo T/A de rs4073 a un efecto protector (OR=0.67 [0.43 a 1.06], *p-value*=0.084). Aunque no se evidencia una asociación estadísticamente significativa, se observan tendencias de ciertos genotipos para tener un efecto protector o de riesgo.

Palabras clave: Estudio de asociación genética, HTLV-1, interleucina 8, Paraparesia espástica tropical, SNP.

## ABSTRACT

Human T-cell lymphotropic Virus Type I (HTLV-1) is a retrovirus that generates an inflammatory response that can trigger a neurological disease called HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP). This inflammatory process is due different cytokines involved in stopping the predominant infiltration of infected CD4<sup>+</sup> T lymphocytes. Interleukin 8 (IL-8) is a pro-inflammatory cytokine that induces the activation and migration of neutrophils towards the sites of infection and it is continuously expressed in HTLV-1 infected cells. The aim of this study was to evaluate the statistical association in a case-control design of two *IL-8* gene polymorphisms, SNPs rs2227543 and rs4073, and the presence of HAM/TSP. The results show that there are not statistically significant differences in the frequencies distribution of genotypes neither haplotypes between the control and the HAM/TSP group. Nevertheless, the genotype T/T from rs2227543 shows a trend for a risk effect (OR=3.40 [0.98 a 11.84], *p-value*=0.048), meanwhile the genotype T/A from rs4073 for a protective effect (OR=0.67 [0.43 a 1.06], *p-value*=0.084). Even if the results do not show a statistically significant association, there is a trend for some genotypes for a protective or risk effect.

Key words: Genetic association studies, HAM/TSP, HTLV-1, Interleukin 8, SNP



## 1. INTRODUCCIÓN

En la actualidad, aproximadamente 10 millones de personas en todo el mundo están infectadas con el virus linfotrópico humano de células T tipo I (HTLV-1) (Gessain y Cassar 2012). De esta población más de un 90% permanece como portadores asintomáticos durante toda su vida, un 4% desarrolla leucemia/linfoma T del adulto (LLTA), un 2% paraparesia espástica tropical (PET) y el resto enfermedades inflamatorias e infecciones oportunistas como por *Strongyloides stercoralis* y tuberculosis (Goncalves *et al.* 2008).

El principal síntoma de PET se caracteriza por una debilidad espástica y progresiva de los miembros inferiores, dolor lumbar y síntomas urinarios que suelen tener una rápida progresión en la edad adulta. Estos síntomas son generados por lesiones inflamatorias focales en el sistema nervioso central (SNC) originadas por la infiltración de linfocitos T CD4<sup>+</sup> y con un ulterior progresivo aumento de linfocitos T CD8<sup>+</sup> (Broglia 2007). Las interleucinas son proteínas involucradas en la regulación de la respuesta inflamatoria que dirige el sistema inmunitario. Entre ellas se encuentra la interleucina 8, codificada por el gen *IL-8*, que tiene una acción quimioatrayente de neutrófilos y linfocitos T (Mori *et al.* 1998).

Por otro lado, se ha evidenciado que polimorfismos de nucleótido simple (SNP) en las regiones promotoras o reguladoras de un gen pueden influir de manera positiva o negativa en la expresión génica (Ramírez-Bello *et al.* 2013). Por tal motivo, la presencia de este tipo de mutaciones puntuales en *IL-8* puede provocar una expresión diferencial y predisponer a una respuesta inflamatoria crónica exacerbada que puede contribuir a la presencia de PET en HTLV-1 positivos.

Los SNPs rs4073 y rs2227543 se ubican en la región promotora y en el intrón 3 del gen *IL-8*, respectivamente. Ambos han sido asociados con distintas enfermedades

inflamatorias, sean de origen sistémico (Suh *et al.* 2011), viral (Hull *et al.* 2000) o bacteriano (Garey *et al.* 2010).

En el Perú, el virus HTLV-1 es considerado endémico para la región y tiene una importante repercusión tanto en la salud como en el aspecto social y económico de las personas. Esta situación es un tema de interés en la salud pública debido a que se han reportado casos de PET en distintas ciudades del país (Gotuzzo *et al.* 2004). Por este motivo, la presente tesis tiene por objetivo evaluar la asociación genética de estos SNPs y PET con la posibilidad de proponer nuevos marcadores genéticos que no solo podrían servir para tomar medidas preventivas frente al desarrollo de la enfermedad, sino también como indicadores para posibles nuevos tratamientos enfocados a la regulación del sistema inmune. En el presente estudio también se ha evaluado el efecto que podrían tener otros factores relacionados al desarrollo de PET, todo esto en favor de entender mejor las bases genéticas de esta patología generada por HTLV-1 y brindar nuevas alternativas a las personas en nuestro país que padecen de una enfermedad que aún no tiene cura.

## 2. MARCO TEÓRICO

### 2.1 Virus HTLV

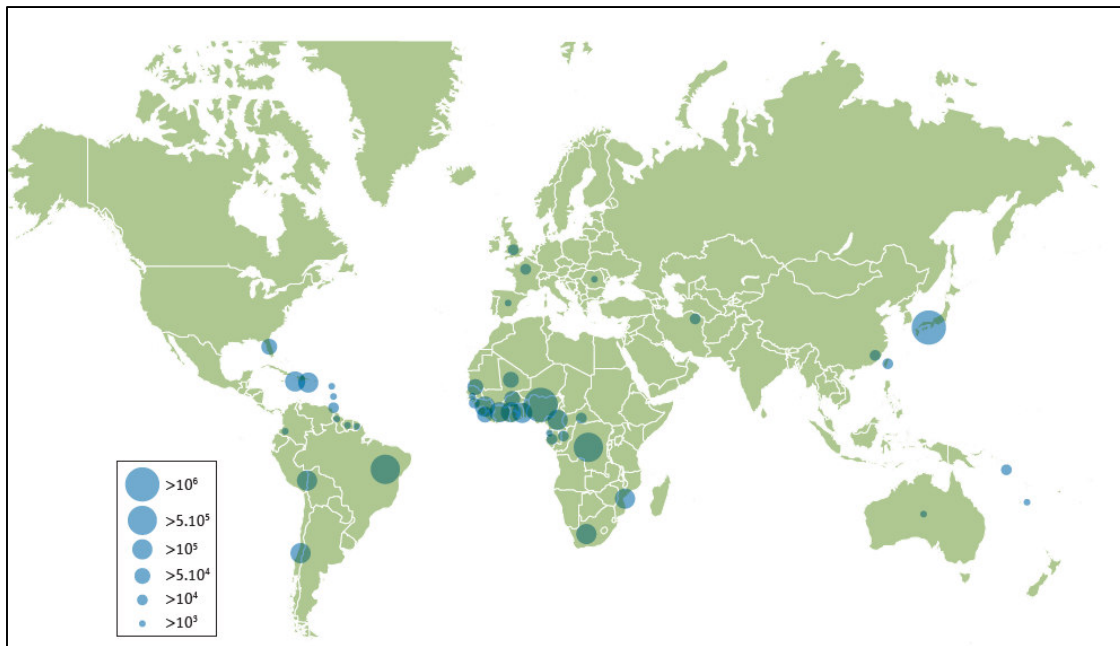
El virus linfotrópico de células T humanas o HTLV (del inglés *Human T-lymphotropic virus*) pertenece a la familia *Retroviridae*, subfamilia *Orthoretrovirinae* y género *Deltaretrovirus*. (De Freitas Carneiro 2010). Hasta ahora se han reportado cuatro tipos de virus HTLV, de los cuales HTLV-1 y HTLV-2 se encuentran en una mayor frecuencia que HTLV-3 y HTLV-4 con solamente cinco casos reportado en África central (Mahieux y Gessain 2011). De ellos, HTLV-1 es el más patogénico.

### 2.2 Epidemiología

HTLV-1 es el primer retrovirus humano reportado (Poiesz et al. 1980). A principios de los años 90 se estimaba alrededor de 10 a 20 millones de personas infectadas con HTLV-1 a nivel mundial (de Thé y Bomford 1993). Sin embargo, debido a que muchos estudios no determinaban la incidencia y prevalencia de la infección, no podían considerarse estudios poblacionales por lo que se necesitaba un reporte epidemiológico más exacto y actualizado (Hlela et al. 2009).

En la actualidad, gracias a los estudios realizados por Gessain y Cassar en el 2012, se estima un número de alrededor de 5 a 10 millones de personas infectadas en todo el mundo. Si bien esta información proviene de estudios realizados en regiones donde ya se conocía de la prevalencia del virus, estos resultados han permitido determinar zonas endémicas a nivel mundial como África, Japón, el Caribe y Sudamérica (Figura 1). En el Perú se ha estimado una cantidad que asciende hasta las 450 000 personas infectadas (Gessain y Cassar 2012). Esto ubica a Perú, dentro de nuestro continente, como uno de los países con mayor endemismo para el virus. Se conocen siete subtipos (A-G) de HTLV-

1, los cuales son reconocidos por la secuencia de su región terminal larga (LTR). El subtipo A es el más cosmopolita de los siete (Cassar et al. 2013).



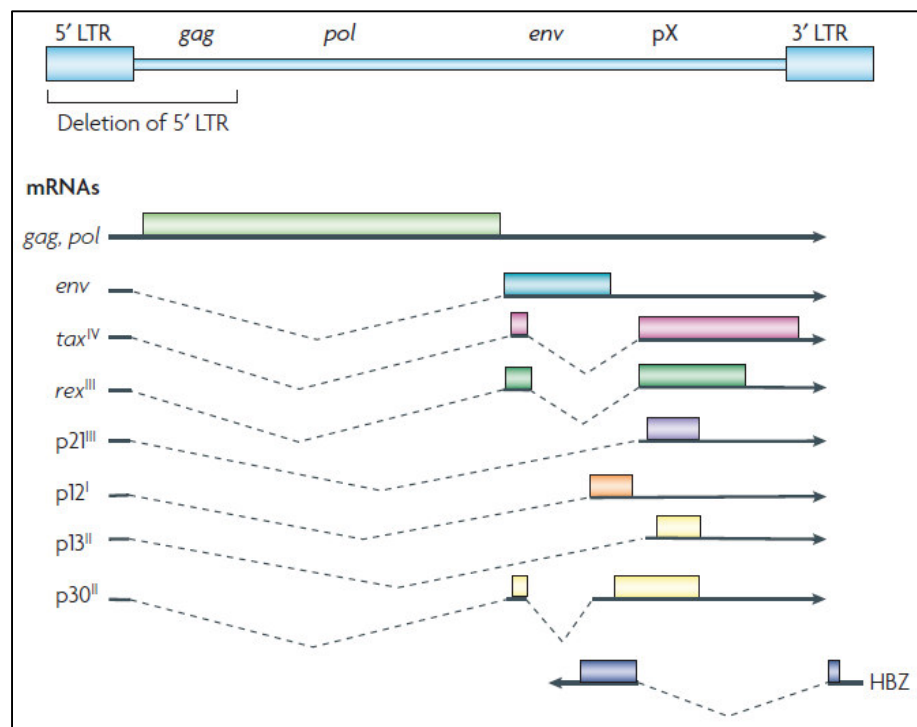
**Figura 1.** Distribución de las regiones endémicas del virus HTLV-1 a nivel mundial (Gessain y Cassar 2012).

### 2.3 Estructura genómica del virus HTLV-1

El genoma de HTLV-1 es de aproximadamente 9kb y contiene genes estructurales (*pol*, *env*, *gag*), genes reguladores (*tax*, *rex*) y genes accesorios (*p12*, *p13*, *p30* y *HBZ*) que están flanqueados por regiones LTR (Figura 2). En la región 5'LTR se localiza el promotor del virus. El gen *pol* codifica la transcriptasa reversa, proteasa e integrasa necesarias para la inserción del virus al genoma del hospedero. El gen *env* se transcribe y traduce en proteínas que residen en la capa lipídica de la envoltura viral, importante para la infección. Asimismo, el gen *gag* codifica las proteínas que forman la cápside viral (Matsuoka y Jeang 2011).

Entre el gen *env* y la región 3'LTR se encuentra la región pX que contiene las secuencias codificantes para p12, p13 y p30, Tax y Rex. Las tres primeras facilitan establecer una

infección viral persistente in vivo evitando la respuesta del sistema inmune (Valeri *et al.* 2010) mientras que Tax y Rex presentan una acción más regulatoria. La proteína Rex modifica un amplio rango de vías para organizar un adecuado ambiente celular que permita la estabilización y traslocación del ARNm viral sin procesar, importante para la autoreplicación efectiva (Nakano y Watanabe 2016). La proteína Tax transactiva la transcripción del mismo provirus y de muchos genes implicados en la activación, proliferación e inhibición de los puntos de control del ciclo celular y reparación de ADN en el hospedero (Bangham *et al.* 2015).



**Figura 2.** Estructura genómica del virus HTLV-1 (Matsuoka y Jeang 2007).

Finalmente, el gen *HBZ* es el único gen del virus codificado en la hebra negativa, actúa para promover la proliferación de células infectadas e inhibir muchas acciones de *tax* (Matsuoka y Yanusaga 2013).

## 2.4 Infección y transmisión de HTLV-1

En seres humanos la transmisión del virus suele darse por tres vías: De madre a hijo por lactancia materna, relaciones sexuales sin protección y transfusiones de sangre. En áreas endémicas de HTLV-1 la transmisión madre-hijo es considerada el factor más importante del mantenimiento endémico (Blattner 1990).

A nivel celular, el virus infecta tanto linfocitos T citotóxicos CD8<sup>+</sup> como T CD4<sup>+</sup> vírgenes, siendo estos últimos capaces de diferenciarse en cuatro líneas mayores: Th1, Th2, Th17 y T regulatorias (Treg). La población celular de linfocitos T CD4<sup>+</sup>CD25<sup>+</sup>CCR4<sup>+</sup> conforman los subgrupos de linfocitos Treg y Th2 en individuos sanos. Son estos dos subgrupos los reservorios virales predominantes para HTLV-1 en pacientes que desarrollan enfermedades asociadas al virus (Araya *et al.* 2011).

## 2.5 Enfermedades asociadas a HTLV-1

Pruebas epidemiológicas corroboran que HTLV-1 está asociado a patologías como: leucemia de linfocitos T adultos (LLTA o ATL, *Adult T-cell Lymphoma*), paraparesia espástica tropical asociada a HTLV (HAM/TSP, *HTLV associated myelopathy/tropical spastic paraparesis*) o simplemente paraparesia espástica tropical (PET), uveítis, dermatitis infectiva e infecciones oportunistas como *Strongyloides stercoralis* o tuberculosis (Verdonck *et al.* 2007). Por otro lado, HTLV-2, 3 y 4 todavía no son auténticamente identificados como la causa de enfermedades en humanos (Ikram *et al.* 2017).

Entender las causas de las enfermedades asociadas a HTLV-1 se hace difícil por la ausencia de un modelo animal adecuado y la inaccesibilidad al tejido del sistema nervioso central (SNC) de individuos infectados (Cook *et al.* 2013).

## 2.6 Paraparesia espástica tropical (PET)

Aproximadamente un 10% de individuos portadores de HTLV-1 desarrolla procesos patológicos que derivan de la infección por HTLV-1; entre ellos, la paraparesia espástica tropical o PET. Esta es una enfermedad sistémica e inflamatoria mediada por el sistema inmune y que puede dañar también tejidos distintos al nervioso, principalmente ocular, reumático y dermatológico (Goncalves *et al.* 2008).

La sintomatología de PET se caracteriza por una debilidad y constante contracción muscular (espasticidad) progresiva de los miembros inferiores, dolor lumbar y síntomas urinarios que suelen tener una rápida progresión en la edad adulta, aunque casos en adolescentes y niños también han sido reportados (Kendall *et al.* 2009). Estos síntomas son generados por lesiones inflamatorias focales del sistema nervioso central originadas por la infiltración de linfocitos T CD4<sup>+</sup> infectados y con un ulterior progresivo aumento de linfocitos T CD8<sup>+</sup> (Broglia 2007). La infiltración de estas células mononucleares promueve una respuesta inflamatoria crónica, produciendo degeneración de mielina y axones en sitios de bajo flujo sanguíneo y en regiones torácicas medianas y bajas de la médula espinal (Izumo 2010). Estos daños generados a nivel del SNC hacen que, con el tiempo, la persona tenga complicaciones al caminar terminando, en algunos casos, postrado en una silla de ruedas. Asimismo, el virus HTLV-1 suele expresar en mayor cantidad la proteína Tax durante los estados de respuesta inflamatoria excesiva que son propios de PET (Saito *et al.* 2009) causando la activación de la vía NF-kB la cual se encuentra incrementada en las personas que presentan PET (Oh *et al.* 2011).

## 2.7 Vía NF-kB en PET

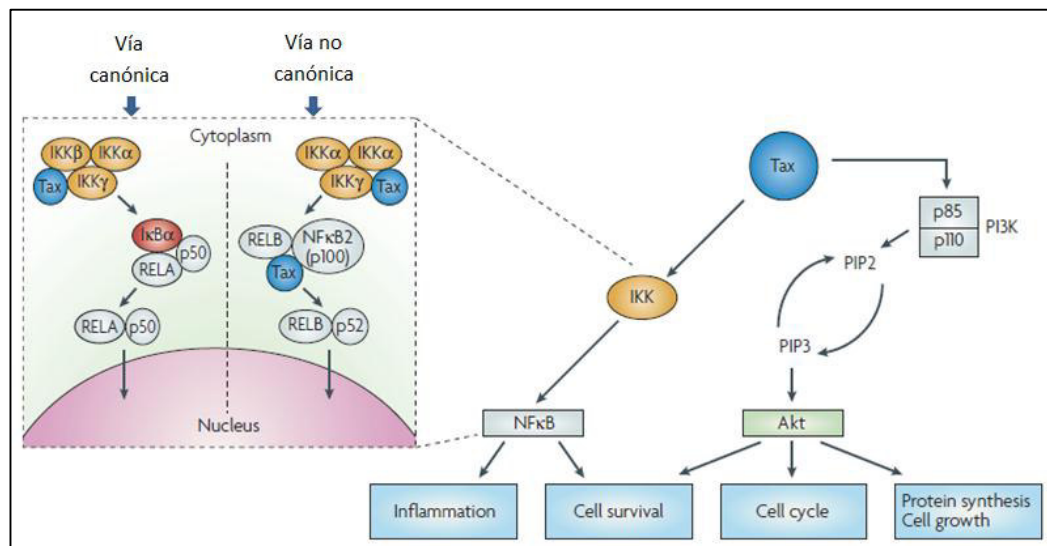
El factor nuclear potenciador de las cadenas ligeras kappa de las células B activadas (NF-kB) es un factor de transcripción que regula la expresión de genes implicados en

supervivencia celular e inflamación. En mamíferos cinco miembros conforman la familia NF- $\kappa$ B: NF- $\kappa$ B1 (p50), NF- $\kappa$ B2 (p52), RelA (p65), RelB y c-Rel.

En células sanas, NF- $\kappa$ B es un dímero que se mantiene en el citoplasma unido a su inhibidor I $\kappa$ B $\alpha$  y que mediante varios estímulos (lipopolisacáridos bacterianos, virus, estrés oxidativo, luz ultravioleta, etc.) es liberado y transportado al núcleo para la transcripción de genes diana (Echeverri y Mockus 2008).

En los linfocitos infectados, la proteína viral Tax lleva a cabo múltiples funciones relacionadas a reorientación de microtúbulos y la activación de las vías CREB, AP-1, Wnt/ $\beta$ catenina, TGF $\beta$ /SMAD y NF- $\kappa$ B (canónica y no canónica).

Tax activa la vía canónica y no canónica de NF- $\kappa$ B (Figura 3). En la vía canónica Tax se une a la quinasa IKK $\gamma$  e inicia la fosforilación de IKK $\alpha$  e IKK $\beta$  para la formación del complejo IKK $\gamma$ -IKK $\alpha$ -IKK $\beta$ . Dicho complejo fosforila a I $\kappa$ B $\alpha$ , lo que conlleva a la degradación de dicho inhibidor a través del proteosoma y la liberación del dímero RelA/p50 para que migre al núcleo y mantenga la respuesta inflamatoria.



**Figura 3.** Activación de la vía canónica y no canónica de NF- $\kappa$ B y activación de la vía Akt mediada por Tax (Matsuoka y Jeang 2007).



En la vía no canónica, NF- $\kappa$ B2, también conocida como p100, se dimeriza con RelB, inhibiendo el transporte de este último al núcleo. Tax se une al dímero y, mediante un proceso de degradación proteosómica, permite la liberación del dímero citoplasmático RelB/p52 hacia el núcleo, sin embargo su acción está mayoritariamente dirigida hacia el control de la maduración de linfocitos B (Peloponese *et al.* 2006).

Dentro de los genes diana de NF- $\kappa$ B se encuentran citocinas, factores de crecimiento, receptores solubles de citocinas, proteínas de respuesta a estrés oxidativo, moléculas de adhesión y moléculas inmunoregulatoras. En el grupo de las citocinas, encontramos a las interleucinas IL-1 $\alpha$ , IL-1 $\beta$ , IL-2, IL-3, IL-6, IL-8, IL-12 y TNF- $\alpha$ , en su mayoría proinflamatorias (Echeverri y Mockus 2008).

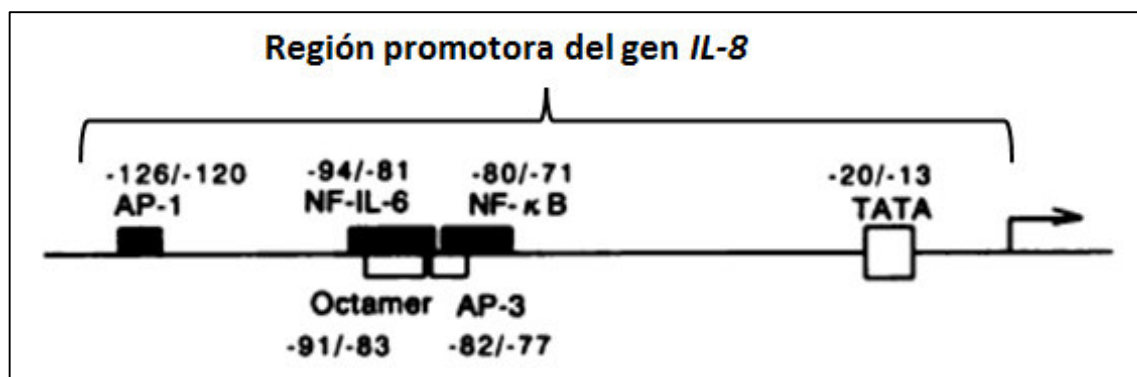
## 2.8 Interleucina 8 y la vía NF- $\kappa$ B

La interleucina 8 es una proteína de tipo quimiocina producida por macrófagos y otros tipos de células como linfocitos T. En humanos, esta proteína es codificada por el gen *CXCL8* o *IL-8* (4q13.3), el cual genera un primer precursor de 99 aminoácidos que luego va a derivar en varias isoformas. Dentro de ellas, la isoforma con una mayor secreción por parte de los macrófagos es la que cuenta con 72 aminoácidos (Brat *et al.* 2005). La proteína IL-8 cuenta con varios receptores de membrana, siendo los receptores acoplados a proteína G, CXCR1 y CXCR2, los más conocidos y expresados generalmente en neutrófilos.

La proteína IL-8 tiene un rol importante en la respuesta inmune innata. Es secretada por macrófagos y monocitos activados para que, mediante su acción quimiotáctica, dirija la migración de neutrófilos, basófilos y linfocitos T hacia las zonas de infección (Rossi y Zlotnik 2000). Asimismo, IL-8 se caracteriza por sus propiedades proangiogénicas, las cuales parecen ser independientes de su actividad quimiotáctica pues se ha observado en ausencia de células inflamatorias (Brat *et al.* 2005).

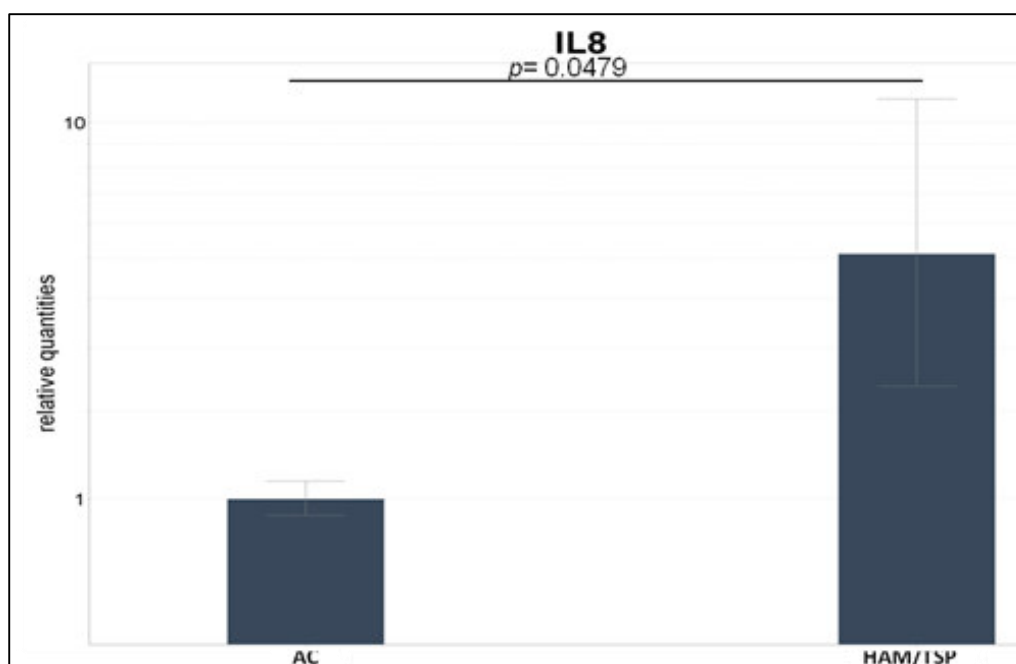
La expresión de *IL-8* en adultos no suele ser constitutiva, sin embargo, en diversas enfermedades inflamatorias se ha evidenciado un aumento en la tasa de transcripción del gen (Grimm *et al.* 1996, Qazi *et al.* 2011). De la misma manera, se ha reportado que células Jurkat (línea celular inmortalizada de linfocitos T) transfectadas con Tax inducen la expresión de *IL-8* y posterior incremento de proteína así como de sus receptores; sugiriendo una acción de quimiocina autocrina a partir de la infección por HTLV-1 (Zhang y Sharma 2000).

A inicios de la década de los 90 se sugirió que la proteína Tax actuaba sobre la expresión del gen *IL-8* a través del factor de transcripción NF- $\kappa$ B. En ese entonces se había descrito un sitio de unión para la subunidad RelA del dímero de NF- $\kappa$ B en la región promotora del gen *IL-8*, lo cual activaba su transcripción (Kunsch y Rosen 1993). Investigaciones posteriores localizan este sitio de unión a RelA en la región -80/-71 *upstream* del gen *IL-8*, así como para otros factores de transcripción como AP-1 (-126/-120) y NF-IL-6 (-94/-81) (Figura 4), sugiriendo que Tax es responsable de la transcripción del gen *IL-8* y de la respuesta inflamatoria que desencadena en PET (Mori *et al.* 1998). La unión de NF- $\kappa$ B a la región promotora y su actividad sobre la transcripción de *IL-8* luego fue corroborada (Bezerri *et al.* 2011).



**Figura 4.** Sitio de unión a NF- $\kappa$ B en la región promotora del gen *IL-8* (Mori *et al.* 1998).

Recientemente se determinaron perfiles de expresión de ARNm de la vía NF- $\kappa$ B en pacientes peruanos con PET. De los 84 genes evaluados, 8 genes mostraron expresión diferencial entre portadores con PET y asintomáticos. Uno de ellos fue el gen *IL-8* que resultó estar sobre expresado en las personas con PET (Talledo 2013) (Figura 5). Esto daría inicio a futuras investigaciones en busca de polimorfismos en el gen *IL-8* que podrían estar afectando dicha expresión exacerbada.



**Figura 5.** Expresión del gen *IL-8* en portadores asintomáticos (AC) y portadores con PET (HAM/TSP) (Talledo 2013).

## 2.9 Factores de riesgo conocidos para el desarrollo de PET

Aparte de la misma presencia del virus, hay otros factores que parecen estimular el desarrollo de la enfermedad. Por ejemplo:

### 2.9.1 Carga proviral

La carga proviral (CPV) mide el número de copias de ADN de HTLV-1 por cada célula mononuclear en sangre periférica (PBMC) tomando en cuenta que cada célula portadora contiene una sola copia proviral (Cook *et al.* 2012). Una vez finalizada la etapa aguda de

infección, el individuo alcanza una carga proviral estable la cual va a depender de las condiciones del hospedero y que tan efectiva sea su respuesta inmune frente a la infección.

Varios estudios reportan una mayor carga proviral en personas con PET con respecto a individuos portadores asintomáticos. Sin embargo, muchos portadores mantienen una carga proviral muy alta incluso antes de que empiecen a desarrollar algún síntoma, por lo que se considera que la carga por sí sola es insuficiente para el desarrollo de PET (Bangham *et al.* 2015).

### 2.9.2 Edad

En general, se sabe que los casos de PET son muy raros en niños. Un estudio realizado a la población de la región endémica de Jamaica y Trinidad y Tobago mostró la más alta incidencia de PET a partir de los cincuenta años (Maloney *et al.* 1998). Esta relación directa entre la edad y número de casos de PET también se refleja en la población peruana donde la edad promedio de inicio de los síntomas es de 45 años (Gotuzzo *et al.* 2004). La influencia de la edad aún no tiene una explicación clara y merece ser evaluada a profundidad pues genera cierta incertidumbre en una investigación de asociación genética al poder describir como paciente asintomático a una persona que posiblemente en unos años desarrolle la enfermedad.

### 2.9.3 Sexo

Según la Organización Mundial de la Salud (OMS), el sexo conforma las características biológicas y fisiológicas que definen a un hombre y a una mujer (OMS 2002). Bajo esa base, los estudios poblacionales realizados en HTLV-1 señalan una mayor incidencia de la enfermedad en mujeres que en varones. Un estudio demográfico en Japón con 383 individuos afectados por PET mostró que cerca del 75% eran mujeres (Coler-Reilly *et al.* 2016), tendencia que también se ha registrado en Trinidad y Tobago, Perú y el Reino

Unido (Martin *et al.* 2010). No obstante, todavía no hay una idea clara de la influencia directa de este factor sobre el desarrollo de la enfermedad.

#### 2.9.4 La genética del hospedero

La poca variación en la secuencia del virus y el hecho de que menos del 5% de todas las personas infectadas presenten PET, sugiere que el genoma del hospedero, y no el genoma viral, sería el que influencia el desarrollo de la enfermedad.

El antígeno leucocitario HLA clase 1 es importante para la especificidad y eficacia de la inmunidad celular mediada por CD8<sup>+</sup>, por lo que también actuaría en el control de la carga proviral. Se ha reportado que el alelo HLA-A\*02 tiene un efecto protector frente al desarrollo de PET tanto en la población de Kagoshima en Japón (Jeffery *et al.* 1999), como en la población de Brasil (Catalan-Soares *et al.* 2009), pero en la población peruana, tras un análisis univariado, no se encontró diferencias estadísticamente significativas entre este alelo y el riesgo a generar PET (Talledo *et al.* 2010).

La presencia de SNPs puede regular la transcripción y la eficiencia de la traducción de ciertos genes. Cuando estos SNPs se encuentran en genes involucrados en la respuesta inmune, pueden afectar la respuesta inmune innata frente a ciertos patógenos y hacer al portador vulnerable a desarrollar enfermedades infecciosas o alérgicas (Liu *et al.* 2012).

En una población de Kagoshima – Japón, se encontró que el SNP IL10-592A localizado en la región promotora del gen *IL10* tiene un efecto protector frente al desarrollo de PET (Sabouri *et al.* 2004). No obstante, dicha asociación no se encuentra en la población de Salvador-Brasil, donde sí se reporta que la citocina pro inflamatoria interleucina 6 presenta un SNP (IL6-634C) que es más frecuente en portadores con PET que asintomáticos (Gadelha *et al.* 2008) lo cual podría ser utilizado como un marcador genético de la enfermedad en esa población.

## 2.10 Análisis de asociación genética

La epidemiología genética es una ciencia que se ocupa de la etiología, distribución y el control de fenotipos de una enfermedad en grupos de parientes, así como las causas hereditarias de estos fenotipos en las poblaciones (Ziegler y König 2014).

Dentro de las herramientas de estudio de la epidemiología genética, los estudios de asociación permiten determinar el efecto de factores genéticos sobre enfermedades y sus posibles interacciones con elementos del medio ambiente (Flores-Alfaro *et al.* 2012). Además, hay distintos diseños de investigación epidemiológica que pueden utilizarse, cada uno con sus respectivas ventajas y limitaciones.

### 2.10.1 Tipos de estudios de asociación genética

#### 2.10.1.1 Diseño caso control

El estudio caso control nos permite evaluar la asociación entre una variable genética y una enfermedad compleja tomando también en cuenta factores de riesgo que podrían influenciar a la aparición de la enfermedad. Dentro de las fortalezas de este diseño es que se puede usar controles no relacionados genéticamente y que representan a la población. Son menos costosos, más rápidos y permiten asociar polimorfismos con la enfermedad de interés (Khoury y Beaty 1996). Sin embargo, se requiere un estricto control de la muestra pues la presencia de estratificación poblacional es uno de los sesgos epidemiológicos más recurrentes que puede afectar en la correcta interpretación de los resultados.

#### 2.10.1.2 Diseño de cohorte

Este es un estudio de seguimiento a una población evaluándose la muestra en el tiempo y recolectando muestras biológicas. Son considerados puntos de partida para diseñar estudio caso-control (Szklo 1998) y nos permite obtener información precisa de la ocurrencia de la enfermedad. Sin embargo, la muestra no es tan representativa de la

población pues los integrantes de la cohorte generalmente pertenecen a un grupo de alta exposición (Kaslow *et al.* 1987), además de tener un mayor costo económico y ser susceptible a cambios metodológicos o de diagnóstico en el tiempo.

#### 2.10.1.3 Diseño basado en familias

Mayormente utilizado para el análisis de enfermedad de herencia mendeliana, pero no es tan apta para enfermedades de rasgos complejos. El hecho de que los integrantes de la muestra pertenezcan a la misma familia, libran al estudio de un posible sesgo epidemiológico por estratificación poblacional. Uno de los problemas con los que cuenta este diseño son los estudios familiares de enfermedades de aparición tardía, donde no se suele encontrar abuelo y nieto o bisnieto al mismo tiempo, lo cual los hace menos eficientes para evaluar polimorfismos de baja eficacia (Flores-Alfaro *et al.* 2012).

#### 2.10.2 Marcadores de ancestría

En los estudios de asociación genética tipo caso control es importante conocer la ancestría individual debido a que la estratificación poblacional en la muestra puede llevar a falsos positivos (Wang *et al.* 2014). Para evidenciar la presencia o ausencia de este sesgo de confusión se utilizan marcadores de ancestría (AIMs, del inglés *ancestry informative markers*): SNPs cuyas frecuencias alélicas permiten diferenciar dos poblaciones distintas. Debido a la gran cantidad de alelos que se pueden obtener al usar varios de estos marcadores por individuo se utiliza un análisis de componentes principales: un algoritmo matemático que reduce la dimensionalidad de la data con una mínima pérdida de información (Ringnér 2008). Estos valores de componentes principales pueden ser representados por *eigenvectors* (eig1, eig2, eig3, etc.) los cuales resumen la información de ancestría en cada individuo (Patterson *et al.* 2006).

### 2.11 SNPs en los estudio de asociación

Los estudios de asociación genética analizan polimorfismos que puedan conferir susceptibilidad a enfermedades complejas. Los polimorfismos genéticos son cambios en la secuencia del genoma presente por lo menos en un 1% entre los individuos de una población. Estas variaciones en la secuencia del ADN pueden ser de distintos tipos: polimorfismos de longitud de fragmentos de restricción (RFLP, del inglés *Restriction Fragment length Polymorphism*), polimorfismos en el número de repetición en tándem (VNTR, del inglés *Variable Number Tandem Repetition*) y los polimorfismos de nucleótido simple (SNPs, del inglés *Single nucleotide polymorphism*).

El polimorfismo de nucleótido simple o SNP es un cambio en un solo nucleótido que puede darse en todo el genoma (regiones codificantes como no codificantes). Los SNPs que están localizados en alta proximidad suelen heredarse en conjunto, siendo esto lo que llamamos desequilibrio de ligamiento o LD (del inglés, *Linkage disequilibrium*) que es la asociación no aleatoria de alelos en diferentes *loci* en una población dada. Esta correlación puede medirse a través de los valores  $D'$  y  $r^2$  con un rango de -1 a 1 y de 0 a 1, respectivamente (Slatkin *et al.* 2008). Cuando se realizan estudios de asociación con dos o más SNPs que se encuentran en LD, se pueden evaluar en conjuntos llamados haplotipos.

El haplotipo es una combinación de alelos de diferentes *loci* en un mismo cromosoma y si bien el análisis individual de SNPs es utilizado ampliamente en los estudios de asociación, el interés por el uso de haplotipos se ha incrementado por su útil enfoque en la localización de genes susceptibles a enfermedades monogénicas (Silverman 2007).

### 2.12 Antecedentes de estudios de asociación de SNPs en *IL-8*

Los SNPs rs4073 y rs2227543, ubicados en la región promotora y en el intrón 3 del gen *IL-8*, respectivamente (Figura 6), han sido ampliamente estudiados en su asociación a



enfermedades inflamatorias. Estos estudios no solo muestran una asociación genética (SNPs – enfermedad) sino también una influencia de estos SNPs sobre el nivel de expresión del gen *IL-8* (Hull *et al.* 2000).

**Figura 6.** Localización de SNPs reportados en el gen *IL-8* en humanos (Ensembl).

El polimorfismo rs2227543 es un SNP (C>T) ubicado en el intrón 3 del gen *IL-8* y también ha sido evaluado por separado y en conjunto con rs4073. Investigaciones de este SNP relacionadas a enfermedades inflamatorias, muestran estudios tipo caso-control en una población de Corea del Sur. Allí se ha encontrado una asociación significativa del SNP

rs2227543 con nefropatía infantil por IgA (Suh *et al.* 2011) y también con bronquiolitis por virus sincitial respiratorio en combinación haplotípica con otros SNPs pertenecientes al mismo gen (Hacking *et al.* 2004).

Los estudios de asociación genética permiten obtener información sobre el posible rol de los SNPs en el desarrollo de una enfermedad. Sin embargo, esta asociación puede variar entre las poblaciones (Suh *et al.* 2011). Por este motivo, es siempre necesario considerar el grupo poblacional como un factor importante para la predicción del riesgo.

Al explicar las posibles causas de asociación entre SNP y enfermedad, se ha evidenciado que la presencia de estos polimorfismos en las regiones promotoras o reguladoras de un gen pueden influir de manera positiva o negativa en la expresión génica (Ramírez-Bello *et al.* 2013). Por tal motivo, la presencia de este tipo de mutaciones puntuales en el gen *IL-8* podría ser la causa de la expresión exacerbada de la interleucina, predisponer a una respuesta inflamatoria crónica y al consecuente desarrollo de PET.

### 2.13 La paraparesia espástica tropical en el Perú.

Se estima que en el Perú hay 450 000 personas infectadas con el virus HTLV-1 (Gessain y Cassar 2012). La mayoría son portadores asintomáticos cuya predisposición a desarrollar una enfermedad asociada al virus es todavía desconocida. De esta población, alrededor del 5% llega a generar enfermedades inflamatorias como PET, ocasionando que la persona, con el tiempo, deje de caminar de manera absoluta. Esto no solo la hace vulnerable a otras infecciones, sino que también aumenta su costo de vida. Además, esta enfermedad, en la actualidad, no cuenta con un tratamiento totalmente efectivo por lo que la propagación del virus puede ser difícil de controlar.

Debido a que la infección por HTLV-1 es frecuente en el Perú el seguimiento médico prioriza la prevención y el tratamiento temprano de las complicaciones infecciosas y el

manejo sintomático de los síndromes inflamatorios (Gotuzzo *et al.* 2010). Aunque el Estado peruano debe otorgar asesoramiento médico gratuito para personas infectadas con HTLV-1 por tratarse también de una infección de transmisión sexual (ITS) (MINSA 2006), hay una insuficiencia en la provisión de tecnologías adecuadas para el diagnóstico y tratamiento de las enfermedades desatendidas, aun cuando el Estado peruano debe asumir su rol subsidiario de acuerdo a lo establecido en la Constitución (Consejo Nacional de Salud 2013).

Investigaciones relacionadas a la transmisión del virus en Guyana francesa reportan que la transmisión de madre a hijo ocurre en el 20% de la descendencia de una madre infectada, lo cual ha sido relacionado a la carga proviral de la mamá, altos título de anticuerpos y periodos de lactancia prolongados (Ureta-Vidal *et al.* 1999). Estos resultados también han sido observados en la población peruana en un estudio que involucra a 343 hijos de madres infectadas. Aquí la frecuencia de HTLV-1 entre los hijos que habían recibido lactancia por menos de 6 meses (4%) era menor que entre los hijos que habían sido lactados por más tiempo (33%) (Gotuzzo *et al.* 2010). En el Perú, estudios realizados sobre prácticas de alimentación y nutrición infantil demuestran que en las tradiciones del ande y la Amazonía el destete se da tardíamente, entre uno y dos años (UNICEF Perú, 2006). En las zonas rurales del país las mujeres suelen tener más hijos, lo que aumentaría la probabilidad de infección y posible desarrollo de la enfermedad si es que la madre fuera portadora del virus. Por ello, medidas de implementación de despistaje gratuito a nivel nacional en mujeres embarazadas podría reducir el riesgo de infección por transmisión vertical y reducir la incidencia de este virus en nuestro país.

#### 2.14 Justificación del estudio

El Perú es considerado una región endémica de HTLV-1 que se mantiene como tal por la alta incidencia del virus y la falta de sintomatología de la mayoría de portadores.

En la actualidad no se cuenta con una cura ni medicación específica para PET. El tratamiento farmacológico solo está dirigido a controlar los síntomas o las causas de la inflamación que en muchos casos no logra revertir el daño neurológico causado al ya diagnosticarse la enfermedad.

Se desconoce todavía mucho sobre la patología. Las bases genéticas y ambientales que hacen prevalecer la enfermedad en solo un pequeño porcentaje de la población aún son exploradas y requieren de atención por el interés científico que representan.

Tomando en cuenta que los estudios de asociación genética de tipo caso-control permiten relacionar un polimorfismo genético con una enfermedad (Lewis y Knight 2009) y que también forman parte de las herramientas usadas en el campo de la farmacogenómica (Ortega y Meyers 2014), la realidad de HTLV-1 en el Perú demanda este tipo de estudios que permitan hallar marcadores genéticos para PET y que puedan ayudar a prevenir o ralentizar los síntomas.

Por este motivo en este trabajo se estudia la asociación entre los SNPs rs4073 y rs2227543 del gen *IL-8* y la presencia de paraparesia espástica tropical en individuos peruanos portadores de HTLV-1. No solamente para evaluar SNPs ya previamente asociados con otras enfermedades inflamatorias, sino por la importancia que supone el entender las bases genéticas del desarrollo de la enfermedad, proponer nuevos posibles tratamientos, descifrar la acción de la respuesta inmune y evitar el deterioro motor indefectible. Esto puede ayudar a mejorar la salud de varios pacientes y reducir el costo social y económico que significa para muchas personas portadoras de HTLV-1.

### 3. HIPÓTESIS Y OBJETIVOS

#### 3.1 Hipótesis

Existe asociación genética entre los SNPs rs4073 y rs2227543 del gen *IL-8* con la presencia de PET en individuos peruanos portadores del virus HTLV-1.

#### 3.2 Objetivos

##### 3.2.1 General.

- Evaluar la asociación genética entre los SNPs rs4073 y rs2227543 del gen *IL-8* con la presencia de paraparesia espástica tropical en individuos portadores del virus HTLV-1 en la población peruana.

##### 3.2.2 Específicos

- Determinar las frecuencias alélicas y genotípicas de los SNPs rs4073 y rs2227543 del gen *IL-8* en la población peruana.
- Evaluar la influencia de los SNPs rs4073 y rs2227543 sobre el desarrollo de la enfermedad con y sin consideración de posibles factores de riesgo.
- Evaluar la influencia de los SNPs rs4073 y rs2227543 en combinación haplotípica sobre el desarrollo de la enfermedad con y sin consideración de posibles factores de riesgo.
- Evaluar la relación de los SNPs rs4073 y rs2227543 con la carga proviral.

## 4. MATERIALES Y MÉTODOS

### 4.1 Declaración ética

El presente proyecto fue aprobado por el comité de ética de la Universidad Peruana Cayetano Heredia (UPCH) en diciembre de 2015, Lima (Perú). Con código SIDISI: 0000066046.

### 4.2 Obtención de la muestra

La muestra consiste en un grupo de individuos peruanos infectados con HTLV-1 que han sido diagnosticados y evaluados en el Instituto de Medicina Tropical “Alexander von Humboldt” (IMTA vH) de la Universidad Peruana Cayetano Heredia (UPCH). El estatus de PET fue determinado por dos expertos de acuerdo a los protocolos internacionales (Osame 1990, De Castro-Costa *et al.* 2006); Se tomaron en cuenta los siguientes criterios de selección:

#### 4.2.1 Criterios de inclusión

a) Grupo de portadores asintomáticos o PA (Control):

- Individuos entre 14 y 88 años.
- Seropositivos para el test de ELISA de HTLV-1
- No presentan ningún daño neurológico o enfermedad asociada a ello.
- Sin relación de parentesco.

b) Grupo de portadores con enfermedad o PET (Casos):

- Individuos entre 14 y 88 años.
- Seropositivos para el test de ELISA de HTLV-1
- Diagnóstico de PET de acuerdo al criterio utilizado en IMTA vH.

- Sin relación de parentesco.

#### 4.2.2 Criterios de exclusión

- Pacientes con diagnóstico clínico no definido.
- Pacientes con enfermedades inflamatorias conocidas.
- Presencia de enfermedades neoplásicas.
- Pacientes relacionados consanguíneamente.
- Pacientes con HTLV-1 pero con otra enfermedad asociada que no sea PET.

#### 4.3 Extracción de ADN

En el estudio se utilizaron muestras de ADN total extraído de sangre de 385 portadores que cumplen con los criterios de selección tanto para el grupo PA como para el grupo PET. Estas muestras ya habían sido usadas en un estudio previo parte del proyecto "*Host genetic immune and viral factors in diseases expression of human T-lymphotropic virus type 1 in Perú*" con código SIDISI: 0000051298. Las muestras de ADN se dividieron en alícuotas de 30ng/μl y se almacenaron a -20°C para evitar su degradación.

#### 4.4 Base de datos

A través del proyecto con código SIDISI: 0000051298, se obtuvo información de cada muestra utilizada en el presente estudio. Dicha base de datos proporcionó información sobre las variables: Sexo, edad, carga proviral y los marcadores de ancestría: eig1, eig2 y eig3.

La variable Sexo está categorizada como: "Masculino" y "Femenino".

La variable Grupos de Edad está categorizada como: "Menos de 45" y "Mayores de 45".

La variable Edad se describe en años.

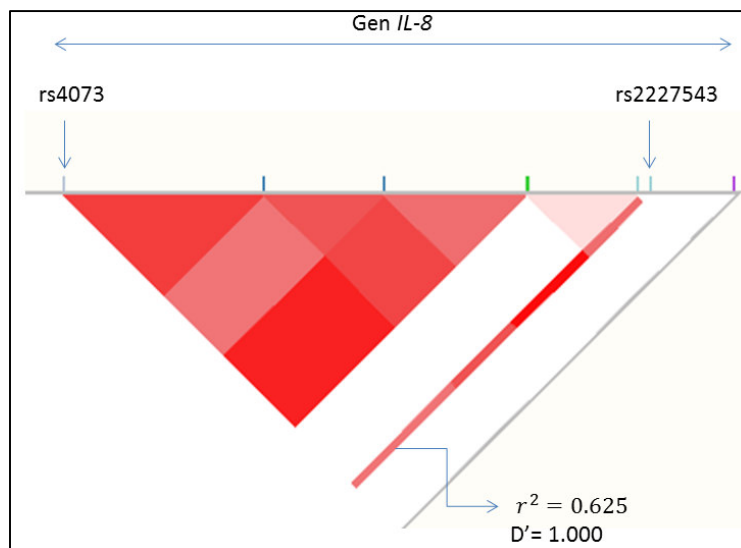
La variable carga proviral (CPV) se describe en número de copias del gen *Tax* por cada 10000 células mononucleares de sangre periférica (PBMC).

Las variables eig1, eig2 y eig3 describen los componentes principales (Anexo 1, 2 y 3)

La carga proviral fue obtenida de acuerdo al método descrito por Talledo *et al.* 2010. Las variables eig<sub>s</sub> se obtuvieron de acuerdo a lo descrito por Talledo *et al.* 2012.

#### 4.5 Selección de SNPs

Los SNPs rs2227543 y rs4073 fueron seleccionados de acuerdo a referencias que reportan su asociación con enfermedades inflamatorias de manera individual (Suh *et al.* 2011) y en combinación haplotípica con otros SNPs pertenecientes al mismo gen (Hacking *et al.* 2004). Haciendo uso de las herramientas del sitio web del proyecto Ensembl (<http://www.ensembl.org/index.html>) (Yates *et al.* 2016) se obtuvo información sobre el desequilibrio de ligamiento entre ambos SNPs a partir de valores estadístico  $D'$  (1.000) y  $r^2$  (0.625) (Figura 7). Indicando que los SNPs se encuentran en desequilibrio de ligamiento y pueden generar haplotipos.



**Figura 7.** Desequilibrio de ligamiento generado con la herramienta web Ensembl (Yates *et al.* 2016)

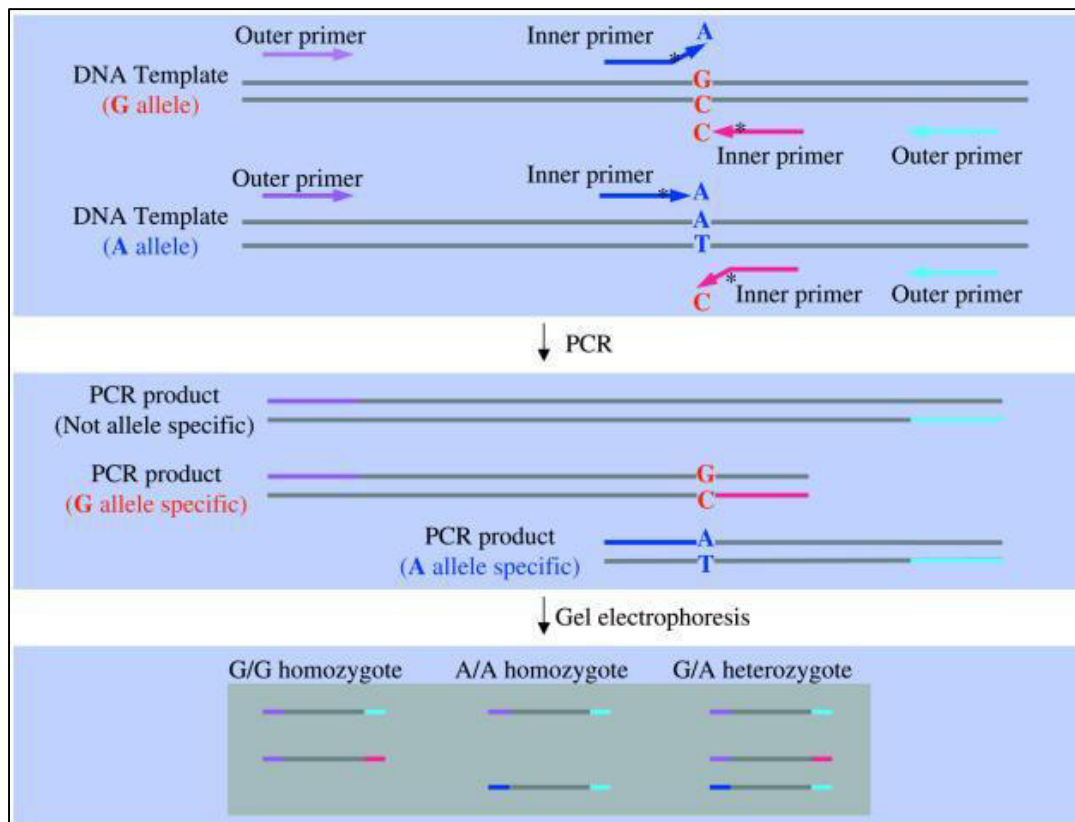


#### 4.6 Genotipificación

La genotipificación se realizó mediante la técnica tetraprimers PCR de sistema de amplificación de mutación refractaria o tetraprimers ARMS-PCR (Figura 8). Dicho sistema requiere el uso de dos pares de cebadores (externos e internos) que van a amplificar los dos alelos diferentes para un SNP en una sola reacción de PCR y al mismo tiempo generar una banda control. El procedimiento, descrito en la Figura 8, se inicia en un SNP (G>A) donde los cebadores *Outer primer* morado (cebador *forward* externo) e *Inner primer* rojo (Cebador *reverse* interno) amplifican el alelo G, mientras que los cebadores *Inner primer* azul (Cebador *forward* interno) y *Outer primer* verde (cebador *reverse* externo) amplificarán el alelo A. Simultáneamente, los cebadores *Forward* externo y *Reverse* externo, al no hibridar específicamente con el SNP, amplificarán independientemente del alelo una banda control de mayor tamaño. Adicionalmente, estos cebadores contienen más de un nucleótido incompatible (*mismatch*) con su región complementaria del extremo 3', lo que aumenta su especificidad (Ye *et al.* 2001).

Dicha genotipificación se realizó en ciego, solo conociendo el código de la muestra mas no si el individuo era PET o asintomático. Tampoco se cuentan con controles positivos debido a que es un análisis exploratorio en la población peruana donde nunca antes se han genotipificado estos SNPs utilizando la metodología tetraprimers ARMS-PCR.

El proceso de genotipificación contó con tres etapas: El diseño bioinformático de los cebadores, la estandarización de los PCRs junto con la amplificación de cada SNP y la electroforesis de los amplicones generados que nos permite saber el genotipo de cada individuo en la muestra.



**Figura 8.** Representación esquemática del método de genotipificación tetraprimer ARMS-PCR (Ye *et al.* 2001)

#### 4.6.1 Diseño de cebadores

Haciendo uso de las herramientas de la web del proyecto Ensembl, se obtuvo la secuencia en formato FASTA de las regiones flanqueantes (*upstream* y *downstream*) del SNP rs2227543 en el genoma humano. Esta secuencia se utilizó para el diseño de los cebadores usando el software PRIMER1 (<http://www.primer1.soton.ac.uk/primer1.html>) (Collins y Ke 2012) elaborado específicamente para el sistema tetraprimers ARMS-PCR. Se indicaron los siguientes parámetros: la posición del SNP en la secuencia, el tamaño de los productos internos y de la diferencia relativa entre ellos. Luego, mediante el programa OligoAnalyzer 3.1 de IDT (<http://www.idtdna.com/Scitools/Scitools.aspx>), se aseguró que los cebadores diseñados tuvieran valores de  $\Delta G$  cercanos a cero para evitar la formación de *hairpins*, *self-dimers* y *hetero-dimers*. También se usó la herramienta mfold

(<http://www.unafold.rna.albany.edu/?q=mfold>) para evaluar las estructuras secundarias que podían formar los amplicones de manera espontánea. Finalmente se utilizó la herramienta BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) para corroborar la presencia de los *mismatch* en los cebadores internos y verificar si hay presencia de posibles productos inespecíficos. Por otro lado, para el SNP rs4073 se utilizaron cebadores para genotipificación por tetraprimers ARMS-PCR previamente reportados (Matheson *et al.* 2006). Los cuatro cebadores generados para cada genotipo se muestran en la Tabla 1.

**Tabla 1.** Secuencias de los cebadores del sistema tetraprimers ARMS-PCR

SNP	Secuencia de cebador	T <sub>m</sub>	Tamaño de fragmento
IL8 T > A rs4073	Cebador <i>forward</i> interno (alelo T)	58 °C	169 pb (alelo T)
	5'-GTTATCTAGAAATAAAAAAGCATACAA		
	Cebador <i>reverse</i> interno (alelo A)		228 pb (alelo A)
	5'-CACAATTTGGTGAATTATCAAA		
	Cebador <i>forward</i> externo		349 pb (banda control)
	5'-CATGATAGCATCTGTAATTAAC TG		
IL8 C > T rs2227543	Cebador <i>reverse</i> externo	59 °C	
	5'-CTCATCTTTTCATTATGTCAGAG		
	Cebador <i>forward</i> interno (alelo C)		216 pb (alelo C)
	5'-TAGCATCAATAGTGAGTTTGTTGTAATC		
	Cebador <i>reverse</i> interno (alelo T)		145 pb (alelo T)
	5'-AACTATGTATGGTCTTTCTGGTCCTA		
	Cebador <i>forward</i> externo		306 pb (banda control)
	5'-GAGACATATAATCCAAAGTCAGCCTATA		
	Cebador <i>reverse</i> externo		
	5'-GACATTCATGTGTTAGAAATGCTCTATA		

#### 4.6.2 PCR

La amplificación de los fragmentos se realizó por PCR convencional usando el termociclador Veriti Thermal Cycle de Applied Biosystems. Para el SNP rs4073, las condiciones del PCR fueron estandarizadas ajustando la concentración del cebador *forward* externo. En caso del SNP rs2227543 se ajustaron las concentraciones de dos cebadores: *reverse* externo y *forward* interno (Tabla 2). En ambos SNPs se usó una reacción de volumen final de 10 µl, sin embargo hay una variación en el protocolo de PCR. Los mejores resultados de amplificación se obtuvieron cuando las temperaturas de hibridación fueron de 54 °C y 58 °C para los SNPs rs2227543 y rs4073, respectivamente (Tabla 3).

**Tabla 2.** Master Mix para PCR de cada SNP

SNPs		rs4073		rs2227543	
Componente	Conc. Inicial	Conc. Final	1Rx (µl)	Conc. final	1 Rx (µl)
H <sub>2</sub> O PCR	(-)	(-)	4.36	(-)	4.26
Buffer	10X	1	1	1	1
MgCl <sub>2</sub>	50 mM	1.5	0.3	1.5	0.3
dNTPs	25 mM	0.2	0.08	0.2	0.08
Cebador F ext	10 µM	0.125	0.125	0.1	0.1
Cebador R ext	10 µM	0.1	0.1	0.125	0.13
Cebador F int	10 µM	1	1	1.1	1.1
Cebador R int	10 µM	1	1	1	1
Enzima	5 U/ µl	0.02	0.04	0.02	0.04
Master Mix			8		8
ADN (30 ng/ µl)			2		2
Volumen final			10		10

Rx: Reacción, Enzima: Taq polimerasa

**Tabla 3.** Protocolo de amplificación de SNPs

SNPs	Denaturación inicial	Temp. denaturación	Temp. Hibridación	Temp. extensión	Extensión final
rs4073	95 °C 10 min	95 °C 1 min	58 °C 1 min	72 °C 1 min	72 °C 5 min
35 ciclos					
rs2227543	95 °C 10 min	95 °C 1 min	54 °C 1 min	72 °C 1 min	72 °C 5 min
35 ciclos					

Temp: Temperatura

#### 4.6.3 Electroforesis

Los productos fueron separados por electroforesis en gel de agarosa al 2% y visualizados por tinción con bromuro de etidio. Para geles de 50 ml se utilizó un voltaje de 90V con una duración aproximada de 30 min. Las imágenes de los geles se revelaron usando el fotodocumentador Molecular Imager® Gel Doc™ de Bio-Rad. Las imágenes fueron editadas con el programa Image lab de Bio-Rad (<http://www.bio-rad.com/es-es/product/image-lab-software>) lo que permitió la identificación de los genotipos de cada individuo.

#### 4.7 Análisis estadístico

Todos los análisis estadísticos: descriptivos, de correlación, de asociación y regresión entre los genotipos y la enfermedad fueron realizados usando la versión del programa estadístico R 3.4.0. Los análisis de asociación (crudo y ajustado) fueron evaluados con el programa SNPstat (<https://www.snpstats.net/snpstats/start.htm?>) (Solé *et al.* 2006) y los gráficos a partir de los resultados obtenidos con el paquete “SNPassoc” de R (<https://CRAN.R-project.org/package=SNPassoc>) (González *et al.* 2014). Todos los test estadísticos presentan un nivel de significación estadística de 0.05 (alfa,  $\alpha = 0.05$ ).

#### 4.7.1 Análisis descriptivo de variables categóricas y numéricas.

Las variables categóricas fueron descritas con las frecuencias absolutas y relativas en cada uno de los grupos. Para la comparativa entre ellos se usaron las pruebas de Chi-cuadrado y Fisher. Esta última prueba se usó siempre que el número de al menos una de las categorías en cualquiera de los grupos sea menor a 5 o por un mensaje de advertencia del programa R.

Las variables numéricas fueron descritas con la media, mediana, desviación estándar y rango. Para la comparativa entre grupos se usó la prueba t o prueba Wilcoxon, ambas para muestras independientes. La elección de una u otra tiene como criterio la distribución (normal o no) de las variables en cada uno de los grupos. Esta evaluación se ha llevado a cabo con la prueba de normalidad de Lilliefors (Kolmogorov-Smirnov) (Gross y Ligges 2015).

#### 4.7.2 Correlación de variables numéricas

La relación entre variables numéricas se evaluó con gráficos de dispersión, líneas suavizadas de tendencia, elipses de correlación y con el coeficiente de relación de Pearson mediante el paquete “Psych” de R (Revelle 2017) siempre que las variables siguiesen una distribución normal.

La variable CPV se transformó con la raíz cuadrada de su valor (rcCPV) con el objetivo de lograr tener una distribución normal (Findley 1990), sin embargo, en el grupo PA, esto no fue posible. No obstante, debido a que su desviación no era muy grande se ha considerado con una distribución normal.

#### 4.7.3 Evaluación de SNPs

La comprobación del equilibrio de Hardy-Weinberg de ambos SNPs se ha usado como detector de posibles errores e la genotipificación o de estratificación poblacional.

Para esto se ha usado la prueba chi-cuadrado con un grado de libertad.

Las frecuencias alélicas (que incluye la frecuencia del alelo menos común o MAF) fueron calculadas a partir de las frecuencias genotípicas. La comparación de las frecuencias genotípicas entre ambos grupos se llevó a cabo mediante la prueba chi-cuadrado o la prueba de Fisher. Adicionalmente se han comparado las frecuencias genotípicas encontradas en este estudio para el grupo PA con las frecuencias reportadas para la población peruana por el proyecto Ensembl.

#### 4.7.4 Análisis de asociación de genotipos

La asociación de uno o más genotipos con la enfermedad se ha estudiado a partir de los valores de *odds ratio* (OR) y sus respectivos intervalos de confianza. Para su obtención se han considerado cinco modelos de herencia que agrupan a los genotipos de acuerdo a su capacidad de modificar el riesgo a generar la enfermedad.

##### 4.8.4.1 Modelos de herencia

- Codominante: Cada genotipo tiene el mismo riesgo a generar la enfermedad.
- Dominante: Poseer una o dos copias del alelo menos frecuente tiene el mismo riesgo.
- Recesivo: Se necesitan dos copias del alelo menos frecuente para modificar el riesgo.
- Sobredominante: Los genotipos homocigotos tienen el mismo riesgo.
- Log-aditivo: Cada copia del alelo menos frecuente modifica el riesgo de forma aditiva.

La elección del modelo que ajuste mejor los datos se determina por el menor valor de AIC (Criterio informativo de Akaike). Los cinco modelos de herencia se analizan bajo tres esquemas generales de regresión logística:

~: “en función de”

#### 1. Crudo (enfermedad ~ SNP)

2. Ajustado 1 (enfermedad ~ SNP + rcCPV)

3. Ajustado 2 (enfermedad ~ SNP + variables numéricas y categóricas)

#### 4.8.4.2 Cálculo de OR

En el análisis crudo los OR se obtienen a partir de las frecuencias absolutas de los genotipos. El genotipo de referencia es el genotipo más frecuente en el grupo control y toma el valor de 1.

**Tabla 4.** Ejemplo de cálculo de OR en un análisis crudo de asociación en el modelo dominante.

Modelo dominante	Control	Paciente	OR
A/A	a	c	1
A/B + B/B	b	d	ad/cb
a, b, c y d: Frecuencias absolutas de los genotipos			

En el análisis ajustado los OR de cada modelo se obtienen a partir de los coeficientes en la ecuación de regresión logística:

$$\text{logit}(p_1) = \alpha + \beta_1 x_1 + \beta_2 x_2$$

Donde,

$p_1$  = probabilidad que ocurra la enfermedad

$\alpha$  = logaritmo de OR cuando  $x_1$  es cero

$\beta_1$  = Coeficiente que representa el logaritmo de OR

$x_1$  = Variable de interés (SNP)

$x_2$  = Covariable (categórica o numérica)

Si se obtiene la función exponencial (antilogaritmo) de  $\beta$  obtenemos el estimado de OR

$$e^{\beta} = OR$$



#### 4.8.4.3 Evaluación de los OR

**Tabla 5.** Efecto del genotipo sobre la enfermedad según sus valores de OR.

OR	Efecto del genotipo
= 1	No asociación
> 1	Asociación, incremento de riesgo
< 1	Asociación, disminución de riesgo o protección

#### 4.7.5 Análisis de asociación de haplotipos

Al igual que los genotipos, la asociación de haplotipos con la enfermedad se ha estudiado a partir de los valores de OR y sus respectivos intervalos de confianza. Las combinaciones haplotípicas de los SNPs rs2227543 y rs4073 son las siguientes: C-T, T-A, C-A y T-T. Además también se han seguido los tres esquemas de asociación: crudo, ajustado 1 y ajustado 2.

#### 4.7.6 Regresión lineal de la carga proviral

El efecto de las diferentes variables sobre la carga proviral (como raíz cuadrada) se evaluó mediante regresión lineal multivariable.

#### 4.7.7 Poder estadístico

Se calculó el poder estadístico de la muestra de estudio a través de la herramienta web Cats, utilizada en estudios de asociación genética (Skol *et al.* 2006). Se tomó en cuenta: el número de casos y controles, una prevalencia de la enfermedad de 5% y la menor frecuencia alélica de los SNPs rs2227543 y rs4073.

## 5. RESULTADOS

El estudio incluye un total de 385 individuos infectados con HTLV-1, los cuales están divididos en los grupos: Individuos portadores asintomáticos (PA) e Individuos portadores con paraparesia espástica tropical (PET). El grupo PA cuenta con 251 individuos (65.2%) y el grupo PET agrupa 134 individuos (34.8%).

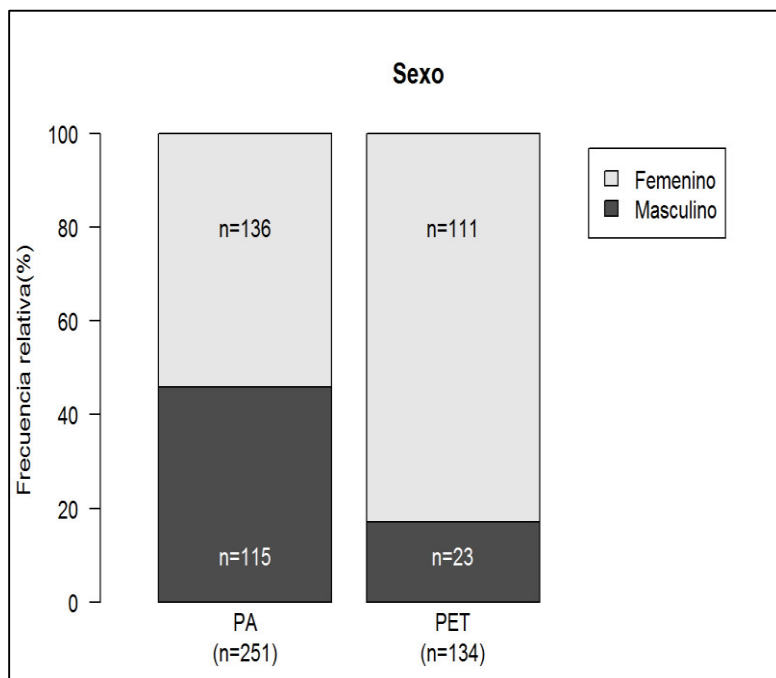
### 5.1 Variables categóricas

**Tabla 6.** Distribución de las variables categóricas en los grupos PA y PET.

Grupos	PA	PET	TOTAL
Variables	n (%)	n (%)	n (%)
Sexo			
Masculino	115 (46)	23 (17)	138 (36)
Femenino	136 (54)	111 (83)	247 (64)
Grupos de Edad			
Menor de 45	127 (51)	36 (27)	163 (43)
Mayor de 45	122 (49)	98 (73)	220 (57)

PA: Individuo asintomático, PET: Individuo con paraparesia espástica tropical, n: Valor absoluto

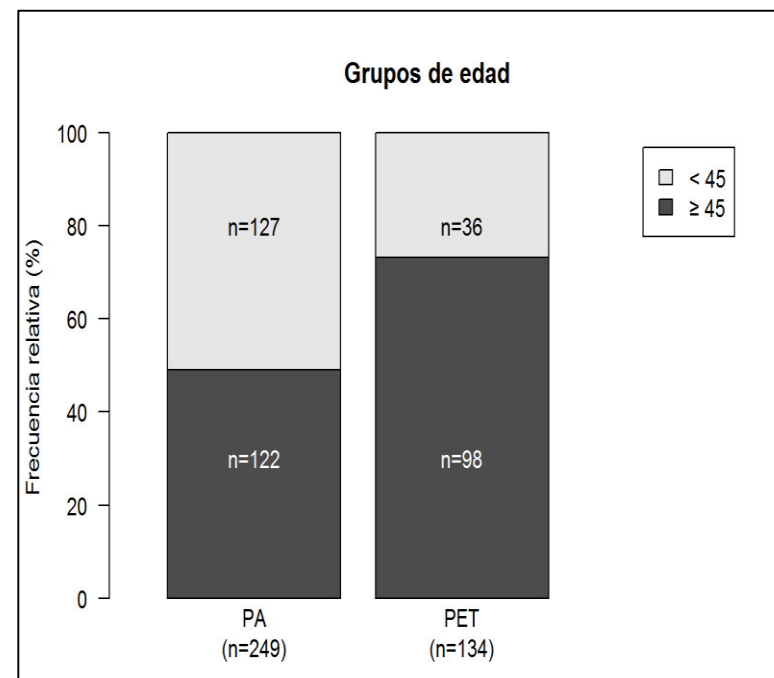
En la variable Sexo se muestra un mayor porcentaje de mujeres dentro del grupo PET (83%), mientras que en la variable Grupos de Edad el mayor porcentaje de personas mayores a 45 años (73%) se ubica también dentro del grupo PET. En la muestra hay dos individuos que no cuentan con datos de edad.



**Test Chi-cuadrado con corrección de Yates**  
 $X^2$ : 29.96  
 gl: 1  
 $p$ -value: 4.42e-08  
 $X^2$ : Chi cuadrado, gl: grados de libertad

**Figura 9.** Frecuencia de individuos por sexo en cada grupo.

Diferencia estadísticamente significativa en la distribución de frecuencias entre PA y PET.



**Test Chi-cuadrado con corrección de Yates**  
 $X^2$ : 19.79  
 gl: 1  
 $p$ -value: 8.652e-06  
 $X^2$ : Chi cuadrado, gl: grados de libertad

**Figura 10.** Frecuencia de individuos por Grupo de edad en cada grupo.

Diferencia estadísticamente significativa en la distribución de frecuencias entre PA y PET.

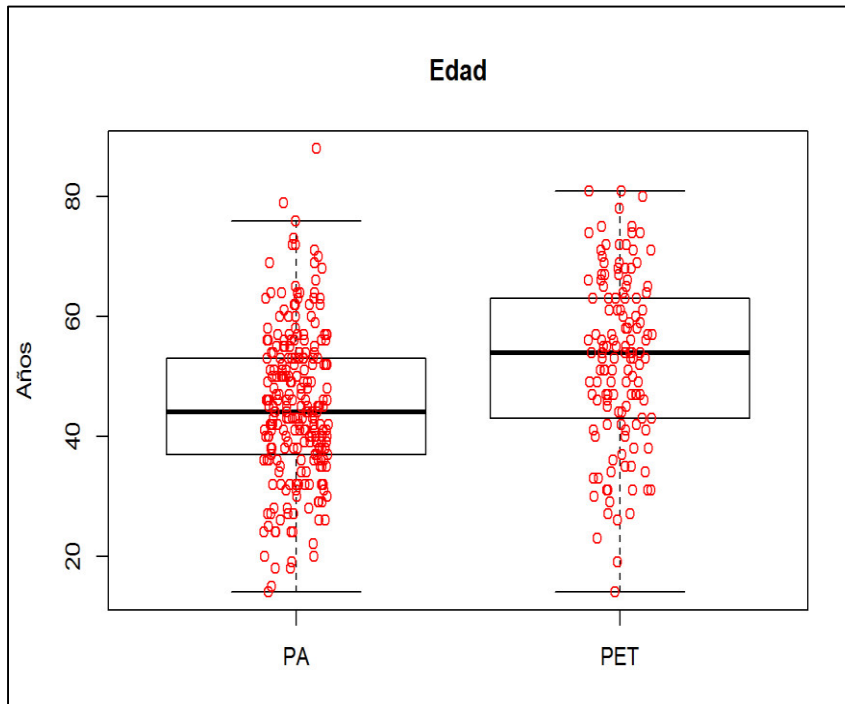
## 5.2 Variables numéricas

**Tabla 7.** Descripción de las variables numéricas en los grupos PA y PET

Grupos	PA						PET					
Variables	min	max	media	mediana	DE	Normalidad <i>p-value</i>	min	max	media	mediana	DE	Normalidad <i>p-value</i>
Edad	14	88	44.99	44	12.52	0.133	14	81	52.68	54	14.18	0.447
CPV	1	9401	1589.5	1070.5	1702.73	<b>1.4e-20</b>	12	9345	3436	2837	2302.39	<b>2.31e-4</b>

min: Mínimo valor, max: Máximo valor, DE: Desviación estándar, Normalidad: Prueba de normalidad Lilliefors (Kolmogorov-Smirnov)

La variable numérica Edad presenta una distribución normal tanto en el grupo PA como en el grupo PET. La variable Carga proviral (CPV) no se encuentra en distribución normal en el grupo PA ni en el grupo PET. En la muestra hay tres individuos que no cuentan con datos de carga proviral.

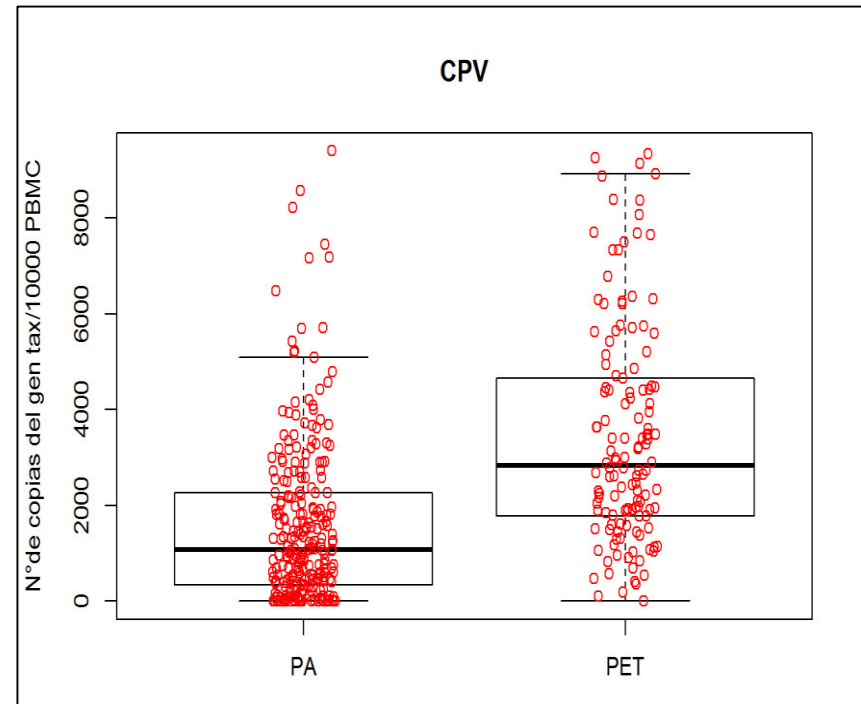


**Prueba t (Welch) para dos muestras independientes**

t: -5.2708  
gl: 244.97  
*p-value*: 2.981e-07

t: Valor de estadístico t, gl: grados de libertad

**Figura 11.** Distribución de la variable CPV en cada grupo. Diferencias estadísticamente significativas en la distribución de Edad entre PA y PET.



**Prueba de suma de rangos de Wilcoxon (muestras independientes)**

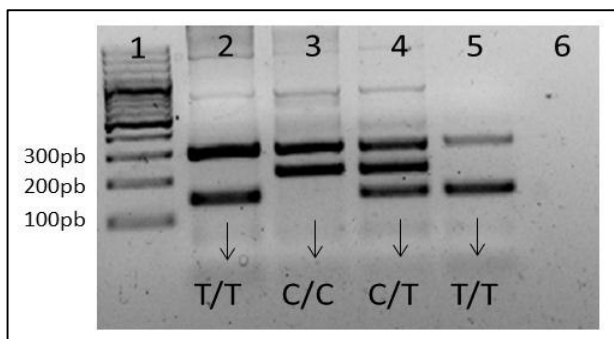
W: 7718  
*p-value*: 5.63e-18

W: Valor obtenido de la prueba de Wilcoxon

**Figura 12.** Distribución de la variable Edad en cada grupo. Diferencias estadísticamente significativas en la distribución de CPV entre PA y PET.

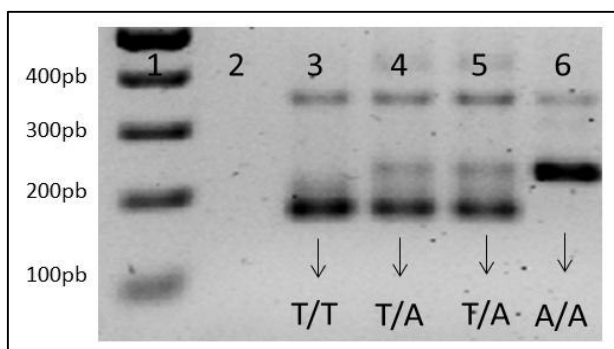
### 5.3 Genotipificación

El método de genotipificación por tetraprimer ARMS-PCR fue exitoso. En ambos SNPs se discrimina claramente la banda control y las de cada alelo. Se realizó la genotipificación de las 385 muestras tras estandarizar el PCR para cada polimorfismo (Tabla 2 y 3). Para verificar la ausencia de contaminación por ADN exógeno, cada placa que contenía las muestras de ADN de los individuos incluía dos pocillos con buffer de elución sin ADN de paciente, lo cual se usó como blanco para la reacción de PCR (Anexo 10).



**Figura 13.** Genotipificación del SNP rs2227543 por método tetraprimer ARMS-PCR.

Electroforesis del SNP rs2227543 del gen IL8 C > T. Banda control (306pb), alelo C (216pb), alelo T (145pb). Carriles: 1 Marcador de ADN (pb); 2, 5 genotipo homocigoto T/T; 3 genotipo homocigoto C/C; 4 genotipo heterocigoto C/T; 6 Blanco.

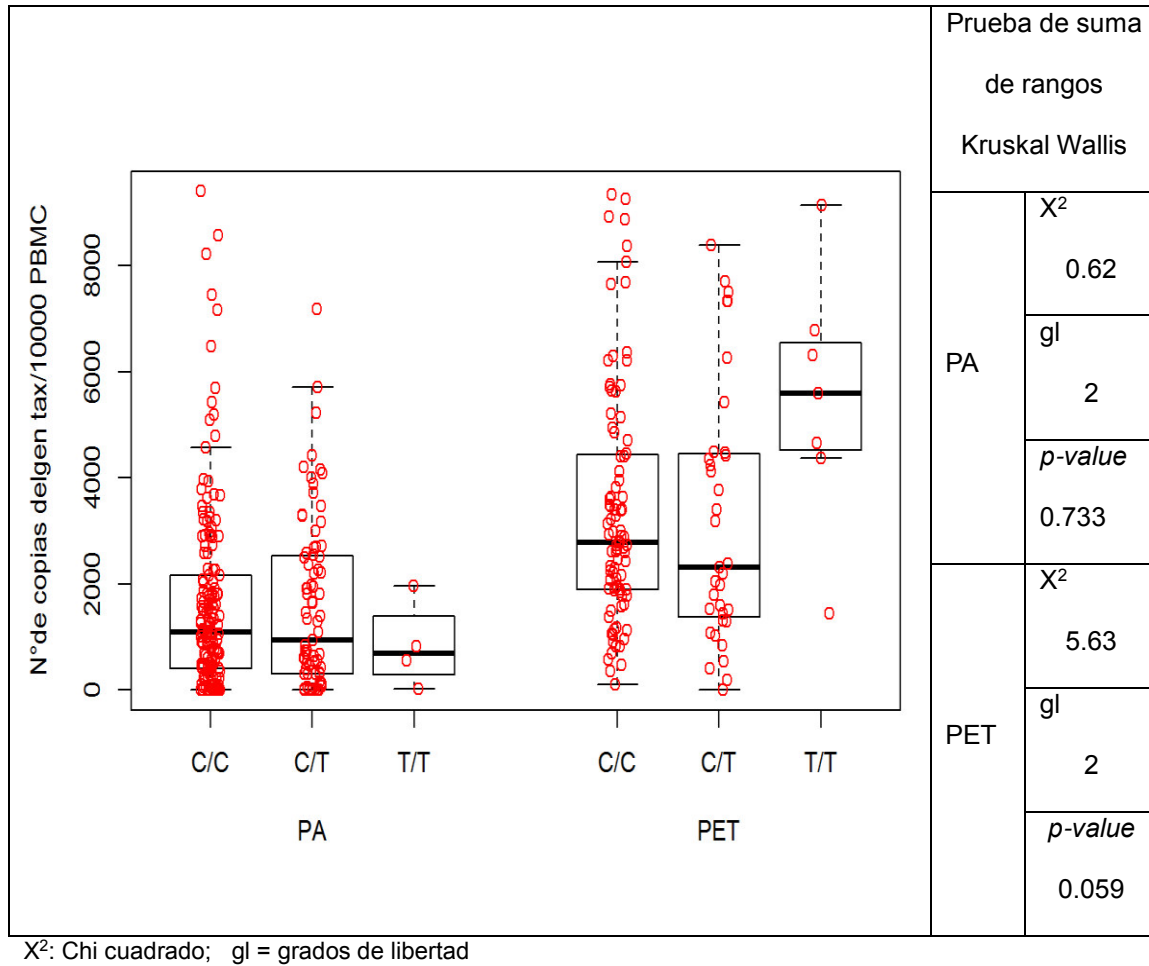


**Figura 14.** Genotipificación del SNP rs4073 por método tetraprimer ARMS-PCR.

Electroforesis del SNP rs4073 del gen IL8 T > A. Bandas control (349pb), alelo A (228pb), alelo T (169pb). Carriles: 1 Marcador de ADN (pb); 2 Blanco; 3 genotipo homocigoto T/T; 4, 5 genotipo heterocigoto T/A; 6 genotipo homocigoto A/A.

## 5.4 Distribución de CPV por genotipos

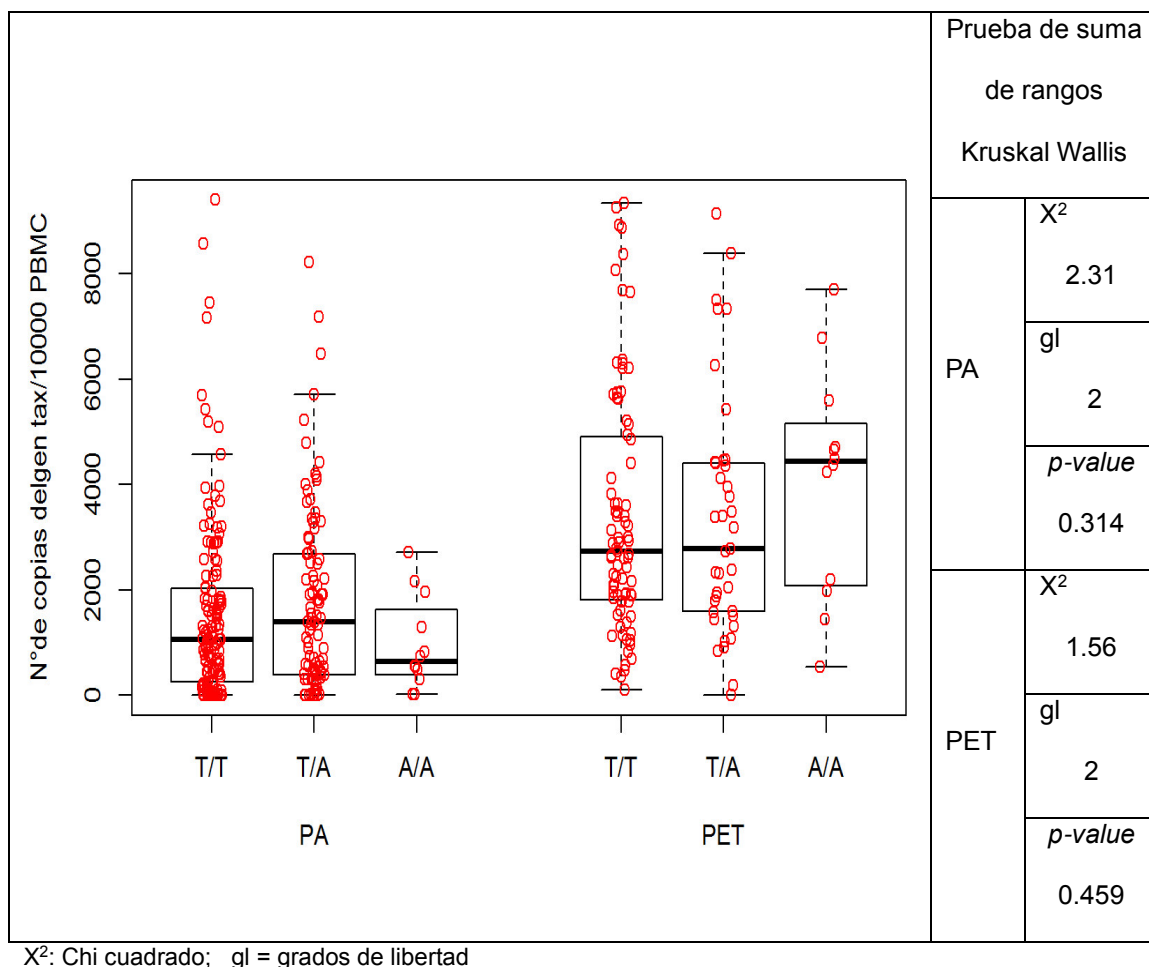
### 5.4.1 SNP rs2227543



**Figura 15.** Distribución de la variable CPV entre los genotipos del SNP rs2227543 en los grupos PA y PET

En general el grupo PET presenta altos valores de CPV en comparación al grupo PA. No se encontraron diferencias estadísticamente significativas entre los genotipos, sin embargo la mediana del genotipo T/T claramente presenta valores muy altos dentro del grupo PET ( > 4000 copias del gen *tax* /10000 PBMC).

#### 5.4.2 SNP rs4073



**Figura 16.** Distribución de la variable CPV entre los genotipos del SNP rs4073 en los grupos PA y PET

Para este SNP tampoco se encontraron diferencias estadísticamente significativas en la distribución de la CPV entre los genotipos de cada grupo. En el grupo PA se aprecia que el genotipo T/A presenta una mediana ligeramente mayor a T/T y A/A, mientras que en el grupo PET la mediana del genotipo A/A presenta un mayor valor de carga proviral en comparación a los otros genotipos.



## 5.5 Análisis de asociación

### 5.5.1 SNP rs2227543

#### 5.5.1.1 Análisis crudo

**Tabla 8.** Análisis crudo de asociación del SNP rs2227543

Asociación del SNP rs2227543 y Enfermedad (N=385, análisis crudo)										
Modelo	Genotipo	PA			PET			OR (IC 95%)	p-value	AIC
		MAF Alelo (fr)	n (%)	EHW p-value	MAF Alelo (fr)	n (%)	EHW p-value			
Codominante	C/C		168 (66.9)	0.18		92 (68.7)	0.15	1	0.096	498.9
	C/T	T (0.17)	79 (31.5)		T (0.18)	35 (26.1)		0.81 (0.50-1.30)		
	T/T		4 (1.6)			7 (5.2)		3.20 (0.91-11.20)		
Dominante	C/C		168 (66.9)			92 (68.7)		1	0.730	501.5
	C/T+T/T		83 (33.1)			42 (31.3)		0.92 (0.59-1.45)		
Recesivo	C/C+C/T		247 (98.4)			127 (94.8)		1	<b>0.048</b>	497.7
	T/T		4 (1.6)			7 (5.2)		3.40 (0.98-11.84)		
Sobredominante	C/C+T/T		172 (68.5)			99 (73.9)		1	0.270	500.4
	C/T		79 (31.5)			35 (26.1)		0.77 (0.48-1.23)		
Log-aditivo	---		---			---		1.07 (0.72-1.58)	0.740	501.5

MAF: Frecuencia del alelo menos común, fr: Frecuencia relativa, N: Total de individuos, n: valor absoluto, EHW: Equilibrio de Hardy-Weinberg, IC: Intervalo de confianza, AIC: Criterio de información de Akaike.

Según los valores de OR y sus respectivos intervalos de confianza, no hay asociación entre ninguno de los genotipos y el estado de enfermedad. El modelo Recesivo llega a ser estadísticamente significativo ( $p\text{-value}=0.048$ ) pero el límite inferior del intervalo de confianza es prácticamente 1 (IC: 0.98 a 11.84). Las frecuencias de MAF son similares en ambos grupos. En las frecuencias genotípicas se observa mayor presencia del genotipo C/C, seguido por C/T y T/T. Los grupos PA y PET se encuentran en EHW.

### 5.5.1.2 Análisis ajustado 1

**Tabla 9.** Análisis ajustado 1 del SNP rs2227543

<b>Asociación del SNP rs2227543 con la Enfermedad (N=382, ajustado por rcCPV*)</b>						
<b>Modelo</b>	<b>Genotipo</b>	<b>PA n (%)</b>	<b>PET n (%)</b>	<b>OR (IC 95%)</b>	<b>p-value</b>	<b>AIC</b>
<b>Codominante</b>	C/C	165 (66.5)	92 (68.7)	1	0.4	421.2
	C/T	79 (31.9)	35 (26.1)	0.84 (0.50-1.42)		
	T/T	4 (1.6)	7 (5.2)	2.26 (0.52-9.89)		
<b>Dominante</b>	C/C	165 (66.5)	92 (68.7)	1	0.73	420.9
	C/T+T/T	83 (33.5)	42 (31.3)	0.92 (0.55-1.52)		
<b>Recesivo</b>	C/C+C/T	244 (98.4)	127 (94.8)	1	0.24	419.6
	T/T	4 (1.6)	7 (5.2)	2.38 (0.55-10.33)		
<b>Sobredominante</b>	C/C+T/T	169 (68.2)	99 (73.9)	1	0.44	420.4
	C/T	79 (31.9)	35 (26.1)	0.81 (0.48-1.37)		
<b>Log-aditivo</b>	---	---	---	1.02 (0.65-1.58)	0.95	421

N: Total de individuos, rcCPV: Raíz cuadrada de carga proviral, n: valor absoluto, EHW: Equilibrio de Hardy-Weinberg AIC: Criterio de información de Akaike.

Los valores de OR y sus respectivos intervalos de confianza indican que no hay asociación entre ninguno de los genotipos y el estado de enfermedad en ninguno de los modelos al ajustar los datos con la variable rcCPV. El tamaño de muestra analizado es de 382 debido a que 3 individuos no cuentan con información de carga proviral.

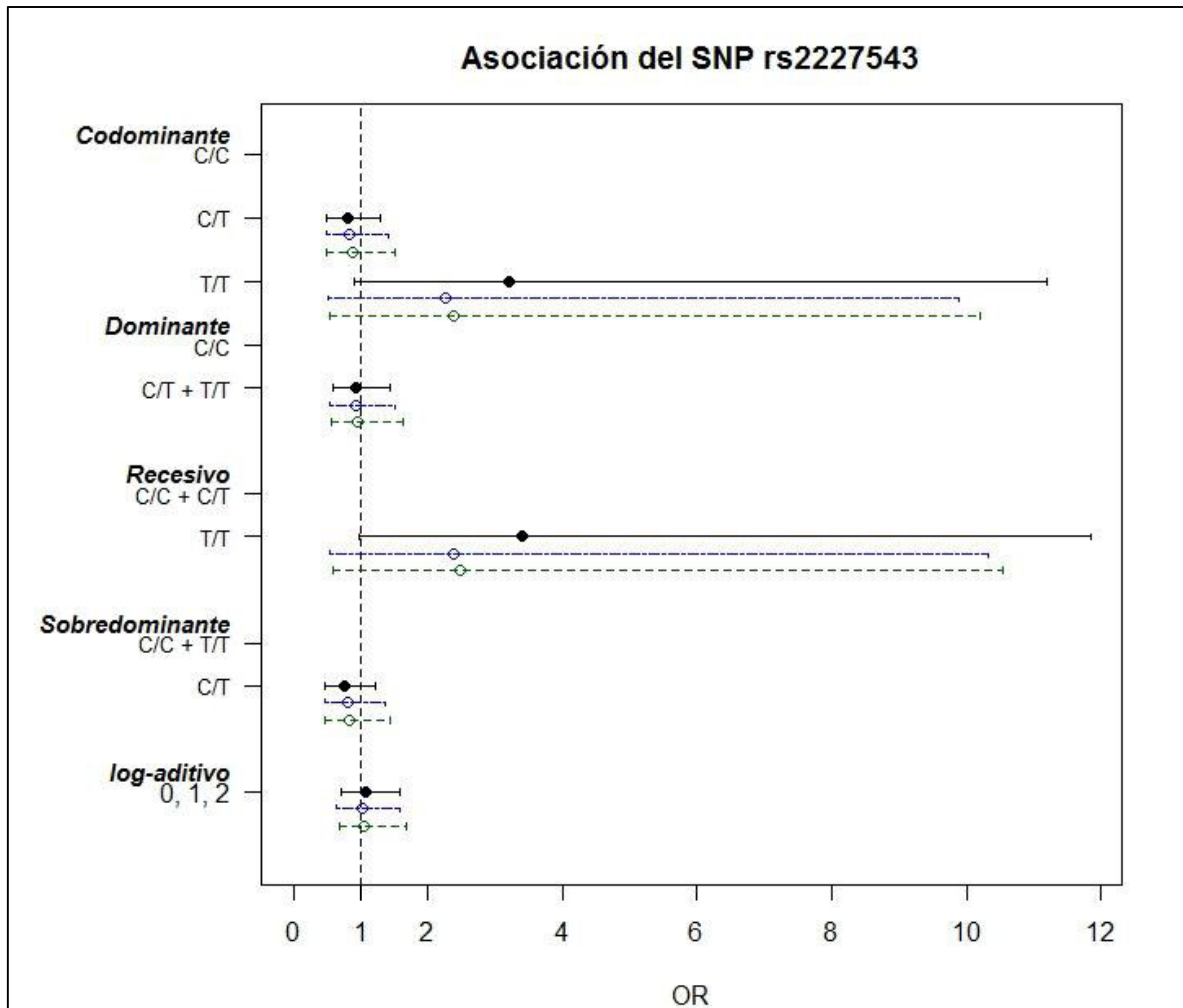
### 5.5.1.3 Análisis ajustado 2

**Tabla 10.** Análisis ajustado 2 del SNP rs2227543

<b>Asociación del SNP rs2227543 con la Enfermedad (N=380, ajustado por variables categóricas y numéricas)</b>						
<b>Modelo</b>	<b>Genotipo</b>	<b>PA n (%)</b>	<b>PET n (%)</b>	<b>OR (IC 95%)</b>	<b>p-value</b>	<b>AIC</b>
<b>Codominante</b>	C/C	163 (66.3)	92 (68.7)	1	0.74	371.7
	C/T	79 (32.1)	35 (26.1)	0.94 (0.52-1.68)		
	T/T	4 (1.6)	7 (5.2)	1.77 (0.37-8.47)		
<b>Dominante</b>	C/C	163 (66.3)	92 (68.7)	1	1	370.3
	C/T+T/T	83 (33.7)	42 (31.3)	1.00 (0.57-1.75)		
<b>Recesivo</b>	C/C+C/T	242 (98.4)	127 (94.8)	1	0.46	369.8
	T/T	4 (1.6)	7 (5.2)	1.80 (0.38-8.56)		
<b>Sobredominante</b>	C/C+T/T	167 (67.9)	99 (73.9)	1	0.77	370.2
	C/T	79 (32.1)	35 (26.1)	0.92 (0.51-1.64)		
<b>Log-aditivo</b>	---	---	---	1.06 (0.65-1.72)	0.82	370.3

N: Total de individuos, rcCPV: Raíz cuadrada de carga proviral, n: valor absoluto, EHW: Equilibrio de Hardy-Weinberg AIC: Criterio de información de Akaike.

Los valores de OR y sus respectivos intervalos de confianza indican que no hay asociación entre ninguno de los genotipos y el estado de enfermedad en ninguno de los modelos al ajustar los datos con las demás variables categóricas y numéricas. El tamaño de muestra analizado es de 380 debido a que 3 individuos no cuentan con información de carga proviral y 2 no cuentan con información de edad.



**Figura 17.** Valores de OR e intervalos de confianza en cada uno de los modelos  
Resumen de los resultados de asociación genotípica del SNP rs2227543 (Tablas 8, 9 y 10) donde cada circunferencia es el valor de OR para cada modelo y las líneas horizontales representan la extensión de los intervalos de confianza.

Análisis crudo ( — ) de color negro.

Análisis ajustado 1 ( - - - - ) de color azul.

Análisis ajustado 2 ( - - - - ) de color verde.

## 5.5.2 SNP rs4073

### 5.5.2.1 Análisis crudo

**Tabla 11. Análisis crudo de asociación del SNP rs4073**

Asociación del SNP rs4073 y Enfermedad (N=385, análisis crudo)										
Modelo	Genotipo	PA			PET			OR (IC 95%)	<i>p-value</i>	AIC
		MAF Alelo (fr)	n (%)	EHW <i>p-value</i>	MAF Alelo (fr)	n (%)	EHW <i>p-value</i>			
<b>Codominante</b>	T/T		144 (57.4)			83 (61.9)		1	0.1	499
	T/A	A (0.24)	95 (37.9)	0.600	A (0.24)	39 (29.1)	<b>0.031</b>	0.71 (0.45-1.13)		
	A/A		12 (4.8)			12 (9)		1.73 (0.75-4.04)		
<b>Dominante</b>	T/T		144 (57.4)			83 (61.9)		1	0.38	500.8
	T/A+A/A		107 (42.6)			51 (38.1)		0.83 (0.54-1.27)		
<b>Recesivo</b>	T/T+T/A		239 (95.2)			122 (91)		1	0.11	499.1
	A/A		12 (4.8)			12 (9)		1.96 (0.85-4.49)		
<b>Sobredominante</b>	T/T+A/A		156 (62.1)			95 (70.9)		1	0.084	498.6
	T/A		95 (37.9)			39 (29.1)		0.67 (0.43-1.06)		
<b>Log-aditivo</b>	---		---			---		0.99 (0.70-1.39)	0.95	501.6

MAF: Frecuencia del alelo menos común, fr: Frecuencia relativa, N: Total de individuos, n: valor absoluto, EHW: Equilibrio de Hardy-Weinberg, IC: Intervalo de confianza, OR: *Odds ratio*, AIC: Criterio de información de Akaike.

Según los valores de OR y sus respectivos intervalos de confianza no hay asociación entre ninguno de los genotipos y el estado de enfermedad. Las frecuencias alélicas son las mismas para ambos grupos (0.24). Al evaluar las frecuencias genotípicas se observa una mayor presencia del genotipo T/T, seguido por T/A y luego A/A. El grupo PET se encuentra en desequilibrio de Hardy-Weinberg, (*p-value*=0.031). El modelo de herencia Sobredominante es el que mejor presenta los datos al poseer el menor valor de AIC.

### 5.5.2.2 Análisis ajustado 1

**Tabla 12.** Análisis ajustado 1 del SNP rs4073

<b>Asociación del SNP rs4073 con la Enfermedad (N=382, ajustado por rcCPV)</b>						
<b>Modelo</b>	<b>Genotipo</b>	<b>PA n (%)</b>	<b>PET n (%)</b>	<b>OR (IC 95%)</b>	<b>p-value</b>	<b>AIC</b>
<b>Codominante</b>	T/T	141 (56.9)	83 (61.9)	1	0.072	417.8
	T/A	95 (38.3)	39 (29.1)	0.62 (0.37-1.04)		
	A/A	12 (4.8)	12 (9)	1.65 (0.63-4.31)		
<b>Dominante</b>	T/T	141 (56.9)	83 (61.9)	1	0.2	419.4
	T/A+A/A	107 (43.1)	51 (38.1)	0.73 (0.45-1.18)		
<b>Recesivo</b>	T/T+T/A	236 (95.2)	122 (91)	1	0.16	419.1
	A/A	12 (4.8)	12 (9)	1.96 (0.76-5.01)		
<b>Sobredominante</b>	T/T+A/A	153 (61.7)	95 (70.9)	1	<b>0.04</b>	<b>416.8</b>
	T/A	95 (38.3)	39 (29.1)	<b>0.59 (0.36-0.98)</b>		
<b>Log-aditivo</b>	---	---	---	0.91 (0.62-1.34)	0.64	420.8

N: Total de individuos, rcCPV: Raíz cuadrada de carga proviral, n: valor absoluto, OR: *Odds ratio* , AIC: Criterio de información de Akaike.

En el modelo Recesivo los valores de OR indican un efecto protector por parte del genotipo T/A (OR: 0.59 e IC: 0.36 A 0.98). Este modelo es el que mejor ajusta los datos (AIC: 461.8) con la variable rcCPV. El tamaño de muestra analizado es de 382 debido a que 3 individuos no cuentan con información de carga proviral.

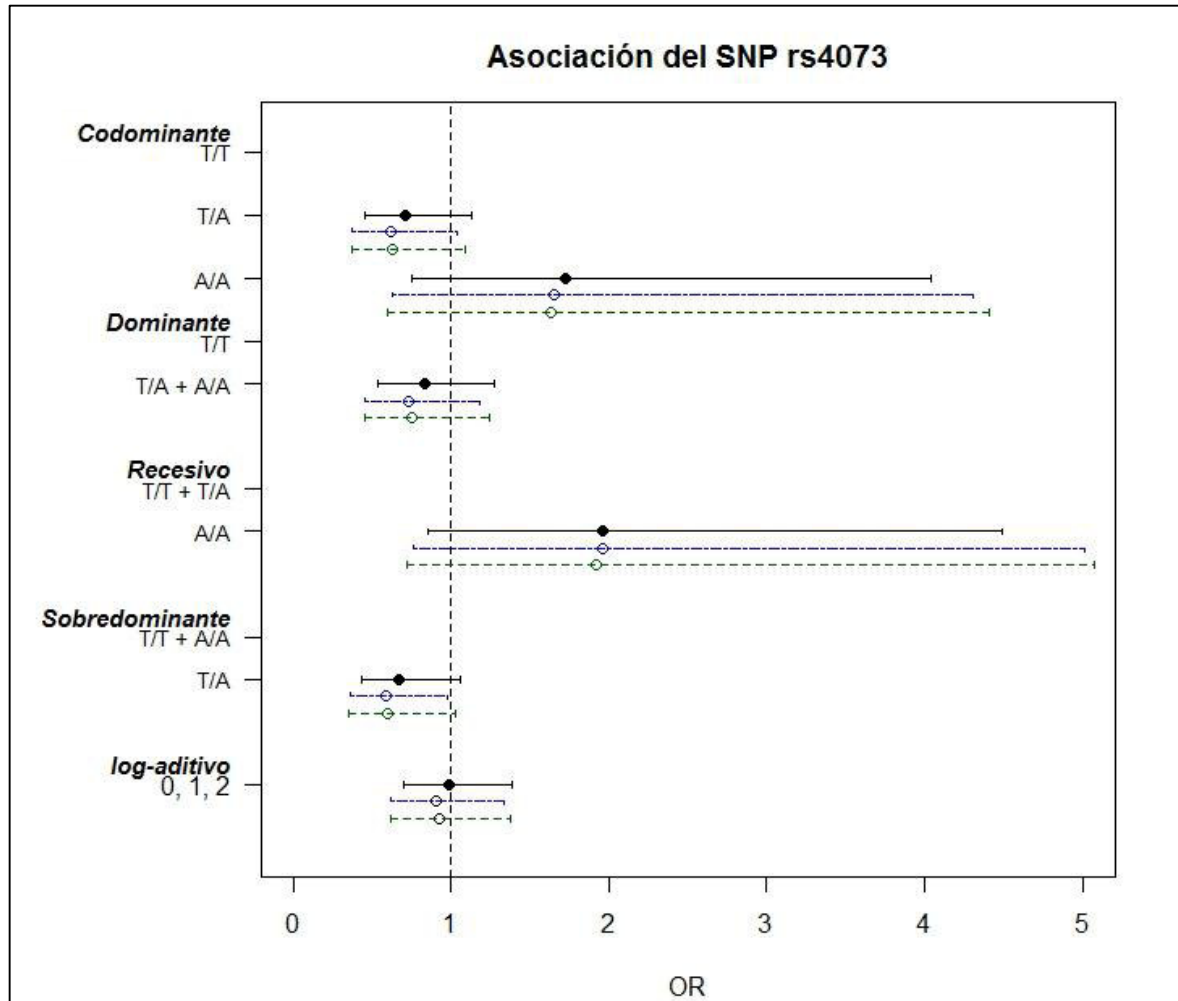
### 5.5.2.3 Análisis ajustado 2

**Tabla 13.** Análisis ajustado 2 del SNP rs4073

Asociación del SNP rs4073 con la Enfermedad (N=380, ajustado por variables categóricas y numéricas)						
Modelo	Genotipo	PA n (%)	PET n (%)	OR (IC 95%)	<i>p-value</i>	AIC
<b>Codominante</b>	T/T	139 (56.5)	83 (61.9)	1	0.1	367.8
	T/A	95 (38.6)	39 (29.1)	0.63 (0.35-1.11)		
	A/A	12 (4.9)	12 (9)	1.76 (0.61-5.05)		
<b>Dominante</b>	T/T	139 (56.5)	83 (61.9)	1	0.28	369.1
	T/A+A/A	107 (43.5)	51 (38.1)	0.75 (0.44-1.27)		
<b>Recesivo</b>	T/T+T/A	234 (95.1)	122 (91)	1	0.17	368.4
	A/A	12 (4.9)	12 (9)	2.07 (0.73-5.84)		
<b>Sobredominante</b>	T/T+A/A	151 (61.4)	95 (70.9)	1	0.064	366.9
	T/A	95 (38.6)	39 (29.1)	0.59 (0.34-1.04)		
<b>Log-aditivo</b>	---	---	---	0.94 (0.61-1.42)	0.75	370.2

N: Total de individuos, rcCPV: Raíz cuadrada de carga proviral, n: valor absoluto, OR: *Odds ratio* , AIC: Criterio de información de Akaike.

Los valores de OR indican no asociación entre ninguno de los genotipos y el estado de enfermedad al ajustar los datos con las demás variables categóricas y numéricas. El tamaño de muestra analizado es de 380 debido a que 3 individuos no cuentan con información de carga proviral y 2 no cuentan con información de edad.



**Figura 18.** Valores de OR e intervalos de confianza en cada uno de los modelos

Resumen de los resultados de asociación genotípica del SNP rs4073 (Tablas 11, 12 y 13) donde cada circunferencia es el valor de OR para cada modelo y las líneas horizontales representan la extensión de los intervalos de confianza. En el modelo Sobredominante se observa los IC del genotipo T/A por debajo de la unidad en el análisis ajustado 2.

Análisis crudo ( — ) de color negro.

Análisis ajustado 1 ( - - - ) de color azul.

Análisis ajustado 2 ( - - - ) de color verde.



## 5.6 Análisis de asociación de haplotipos

No se encontró asociación entre ninguno de los haplotipos y la presencia de PET.

### 5.6.1 Análisis crudo

**Tabla 14.** Análisis crudo de haplotipos

Asociación de haplotipos con Enfermedad (N=385, análisis crudo)					
Nº	rs2227543	rs4073	Frecuencia	OR (IC 95%)	<i>p-value</i>
1	C	T	0.7336	1	---
2	T	A	0.1466	1.09 (0.71 - 1.66)	0.7
3	C	A	0.0897	0.84 (0.48 - 1.44)	0.52
4	T	T	0.03	0.90 (0.37 - 2.22)	0.82
Asociación haplotípica global <i>p-value</i> : 0.89					
N: Total de individuos, OR: <i>Odds ratio</i>					

El haplotipo más frecuente es C – T (0.7336), seguido por T – A (0.1466), C – A (0.0897) y T – T (0.03). En el análisis crudo ningún haplotipo está en asociación con la enfermedad (*p-value*=0.89).

### 5.6.2 Análisis ajustado 1

**Tabla 15.** Análisis ajustado 1 de haplotipos

<b>Asociación de haplotipos con Enfermedad (N=382, ajustado por rcCPV)</b>					
Nº	rs2227543	rs4073	Frecuencia	OR (IC 95%)	<i>p-value</i>
1	C	T	0.7316	1	---
2	T	A	0.1478	0.96 (0.60 - 1.56)	0.88
3	C	A	0.0904	0.85 (0.47 - 1.55)	0.6
4	T	T	0.0302	1.23 (0.43 - 3.59)	0.7
Asociación haplotípica global <i>p-value</i> : 0.92					

Al ajustar los datos por la rcCPV se pierde la participación de 3 individuos que no cuentan con valores de carga proviral (N=382). No hay asociación con ningún haplotipo.

### 5.6.3 Análisis ajustado 2

**Tabla 16.** Análisis ajustado 2 de haplotipos

<b>Asociación de haplotipos con Enfermedad (N=380, ajustado por variables categóricas y numéricas)</b>					
Nº	rs2227543	rs4073	Frecuencia	OR (IC 95%)	<i>p-value</i>
1	C	T	0.7302	1	---
2	T	A	0.1486	0.95 (0.56 - 1.62)	0.86
3	C	A	0.0908	0.96 (0.51 - 1.82)	0.9
4	T	T	0.0303	1.79 (0.56 - 5.71)	0.33
Asociación haplotípica global <i>p-value</i> : 0.79					

Al ajustar los datos con las demás variables categóricas o numérica (Edad, Sexo, etc.) se pierde la participación de 5 individuos en el estudio (N=380). No se observa ninguna asociación con la enfermedad.

### 5.7 Regresión lineal con la variable CPV como respuesta

Se usa la variable CPV transformada (rcCPV) para evaluar su relación con las variables independientes.

**Tabla 17.** Análisis de regresión lineal

rcCPV	Coef.	t	P< t	IC 95%	
rs2227543 C/T	-4.88	-1.64	0.103	-10.74	0.98
rs2227543 T/T	8.22	1.1	0.273	-6.51	22.96
rs4073 T/A	5.53	1.95	0.0526	-0.062	11.11
rs4073 A/A	-0.41	-0.075	0.94	-11.14	10.32
Grupo: PET	21.56	8.85	<b>2e-16</b>	16.77	26.35
Sexo: Femenino	-3.47	-1.5	0.135	-8.025	1.08
Edad	-0.052	-0.4	0.692	-0.31	0.2
eig1	17.52	1.15	0.253	-12.55	47.59
eig2	23.4	1.5	0.135	-7.33	54.12
eig3	15.046	0.98	0.328	-15.17	45.26
Grupo Edad: Menor de 45	-0.46	-0.13	0.898	-7.45	6.53

Coef: Coeficiente de relación, t: Valor del estadístico t, P<|t|: p-value, IC: Intervalo de confianza

La variable Grupo es la única que muestra tener un efecto estadísticamente significativo sobre con la carga proviral (*p-value*: 2e-16).

## 6. DISCUSIÓN

Muchos estudios han descrito a PET en el contexto de enfermedades autoinmunes generadas por HTLV-1, y a pesar de que aún está en debate, la alteración de interleucinas promovida por este virus está críticamente envuelta en desequilibrar el balance entre la respuesta inflamatoria y anti-inflamatoria (Quaresma *et al.* 2015). Debido a que existe evidencia de la relación entre la interleucina IL-8 sobre el desarrollo de PET, la evaluación de polimorfismos dentro de la secuencia cercana al promotor así como en regiones intrónicas, son de gran importancia ya que podrían tener un efecto en la actividad proteica debido a la expresión génica o a su estructura, además de servir como marcadores de la enfermedad.

### *6.1 Frecuencias alélicas y genotípicas*

Los SNPs rs4073 y rs2227543 del gen *IL-8* ya han sido previamente evaluados en distintas poblaciones con respecto a su asociación con diversas enfermedades inflamatorias con origen viral y bacteriano (Hull *et al.* 2000, Garey *et al.* 2010). Es importante observar que las frecuencias alélicas y genotípicas de estos SNPs varían entre las poblaciones según el grado de relación entre estas. Por ejemplo, se han reportado las frecuencias alélicas del SNP rs4073 en poblaciones asiáticas y europeas donde el alelo A es el menos frecuente, caso contrario a lo que ocurre en países del continente africano, donde es el alelo T el menos frecuente (Savage *et al.* 2006). Es posible que la presencia española durante más de 300 años en territorio peruano haya influenciado la frecuencia alélica de este SNP, pues se ha reportado en poblaciones Ashaninkas, Shimma y Quechua (poblaciones nativas peruanas) frecuencias mucho menores a las europeas (Zamudio *et al.* 2016) pero similares a las de este estudio que solo comprende portadores peruanos.

Debido a la variabilidad genética en distintas poblaciones a la nuestra es importante recordar que las frecuencias alélicas y genotípicas obtenidas en este estudio pueden ser diferentes a lo reportado en otras investigaciones. Sin embargo, deben seguir siendo representativas de la población peruana.

La razón de la evaluación del equilibrio de Hardy-Weinberg (EHW) en el grupo control se debe a que este resultado se puede considerar como un control de calidad del proceso de genotipificación. Además, la salida del equilibrio puede ser un indicador de estratificación poblacional. No obstante, cabe resaltar que el grupo PET del SNP rs4073 se encuentra en desequilibrio de Hardy-Weinberg ( $p\text{-value}=0.031$ ). Siendo este el grupo de casos, no invalida el estudio de asociación, más bien da indicios de que un factor disruptivo estaría afectando la distribución esperada de las frecuencias de los genotipos.

La concordancia entre los resultados del presente estudio y los del proyecto Ensembl (Anexo 7 y 9) apoyan el proceso de genotipificación. Además, ambos resultados, en conjunto, permiten estimar de manera más precisa los valores poblacionales de las frecuencias alélicas y genotípicas.

### *6.2 Evaluación de los factores de riesgo*

Respecto a la variable sexo, la presencia mayoritaria de mujeres en el grupo PET es acorde con los reportes epidemiológicos de distintos países como lo descrito para Inglaterra donde un 79.2% de pacientes con PET eran mujeres (Martin *et al.* 2010) o en la población japonesa con un 74.2% (Coler-Reilly *et al.* 2016). En el caso de Perú se han reportado casos de PET con una proporción mujer:hombre de 1.9:1, siendo la muestra proveniente de distintas regiones geográficas del país (De las Casas *et al.* 1996). Los resultados en la muestra de este estudio indican que un 82.8% de casos PET son de sexo

femenino, por lo que la distribución de las frecuencias de sexo es estadísticamente significativa entre PA y PET (Figura 9).

La variable categórica Grupo de Edad muestra que un 73.1 % de personas mayores de 45 años se encuentran en el grupo PET (Figura 10). El incremento de la prevalencia con la edad ya ha sido anteriormente puesto en evidencia como factor importante para el desarrollo de la enfermedad. Esta asociación se explica con la acumulación de seroconversión para HTLV-1 en el curso de la vida y el incremento en el número de parejas sexuales (Villarde *et al.* 2011). De la misma manera, al analizar la edad como variable numérica, se observa también una diferencia estadísticamente significativa en su distribución entre los grupos PA y PET (Figura 11). Esto refuerza los resultados obtenidos por la forma categórica de la variable, donde también se aprecia que a mayor edad, mayor prevalencia de la enfermedad.

La carga proviral es uno de los factores de riesgo más utilizados para predecir la presencia de PET. Investigaciones al respecto han reportado valores de 114 copias por 10000 células, con 78.2% de sensibilidad, para identificar verdaderos pacientes con PET (Furtado *et al.* 2012). El presente estudio muestra valores de carga proviral muy por encima de las 1000 copias (Figura 12), con lo que se evidencia de manera definitiva la condición de enfermedad en nuestro grupo de casos.

El carácter exploratorio de este estudio evidencia que en la población peruana las variables Sexo, Edad y Carga proviral mantienen su poder como factores de riesgo de PET y por ello son utilizadas en los análisis de asociación 1 y 2 de este estudio.

### *6.3 Análisis de asociación de genotipos*

Aunque no se encuentran resultados estadísticamente significativos para la asociación de genotipos y el desarrollo de PET, es importante considerar las tendencias de efecto por

parte de algunos genotipos. En el SNP rs2227543 se observa un efecto de riesgo por parte del genotipo T/T bajo los modelos codominante y recesivo, no obstante, cuando se realizan los modelos ajustados, este efecto se pierde. Además, el tamaño del intervalo de confianza supone una incertidumbre grande con respecto al valor poblacional de riesgo (OR poblacional). La muy poca cantidad de individuos que presentan el genotipo T/T, aumenta la amplitud del intervalo de confianza.

A pesar de su localización intrónica, el SNP rs2227543 ha sido ampliamente evaluado en distintas enfermedades involucradas en la regulación del sistema inmune. Por ejemplo, en la población China, no se encontró una asociación entre este SNP y el riesgo a anemia aplásica adquirida (Zhang *et al.* 2017), pero, en una población de niños coreanos, sí se encontró que el genotipo C/T representaba un factor de riesgo al desarrollo de nefropatía infantil IgA con un OR de 1.64 (IC 95% 1.12 a 2.40) (Suh *et al.* 2011). En el presente estudio es el genotipo T/T quien aparenta tener una acción de riesgo al desarrollo de PET y dicho resultado es acorde con otro realizado en el Reino Unido, donde se evidenció la asociación del alelo T en un haplotipo con la susceptibilidad a desarrollar bronquiolitis por virus sincitial respiratorio (Hacking *et al.* 2004).

Un acercamiento importante en el análisis de este SNP se realizó en un estudio de asociación con la enfermedad de Behcet y la enfermedad de Vogt Koyanagi Harada que muestran un cuadro predominante de uveítis, enfermedad que también deriva de la infección por HTLV-1 (Kamoi y Mochizuki 2012). En los resultados se aprecia que el genotipo C/T tiene un OR de 0.75 (IC 95% 0.0940 a 0.552) mostrando una respuesta de protección a la enfermedad. Sin embargo, dicho resultado fue desestimado tras la corrección por comparaciones múltiples, con lo que se perdió la significancia estadística (Huang *et al.* 2017).

Los SNPs localizados en regiones intrónicas como este pueden afectar la expresión de genes, sea porque su ubicación coincide con las regiones potenciales de empalme alternativo del ARNm o porque se encuentran en desequilibrio de ligamiento con los SNPs que sí son funcionales, así como sucede en el SNP rs2665802 del gen de la hormona de crecimiento humano I (Millar *et al.* 2010) ubicado en el intrón 4 del gen y asociado a bajos niveles de hormona circulante. No obstante, en mi conocimiento, al no existir estudios previos que asocien a rs2227543 con un aumento en la carga proviral de HTLV-1 u otro virus es complicado explicar la posible acción de este SNP intrónico y cómo, de manera particular, podría aumentar la presencia del virus en un individuo portador.

En el SNP rs4073 el análisis crudo no muestra asociación de ningún genotipo con el estado de enfermedad. Sin embargo, se observa una tendencia del genotipo T/A a un efecto protector y del genotipo A/A a un efecto de riesgo (Figura 18). A diferencia del SNP rs2227543, en donde la tendencia a un efecto de riesgo por parte del genotipo T/T se pierde, los modelos ajustados 1 y 2 para el SNP rs4073 no se diferencian notablemente del modelo crudo. Los valores de OR no presentan un cambio notorio por lo que parece lógico atribuir este efecto a los genotipos. En el caso del genotipo T/A también se debe resaltar que el intervalo de confianza implica una menor incertidumbre y una mejor precisión. Así, la significación estadística obtenida en el ajuste 1 es acorde con la idea, mencionada anteriormente, de que el genotipo T/A tendría un efecto protector, como ya se había reportado para otras enfermedades inflamatorias.

Al analizar la asociación del SNP rs4073 con enfermedades autoinmunes generadas por virus, se encuentra un estudio en la población del Reino Unido, el cual muestra al genotipo A/A como el de riesgo al desarrollo de bronquiolitis por virus sincitial respiratorio. Asimismo señalan un incremento en la producción de la proteína IL-8 ante la presencia del genotipo A/A respecto a T/A y T/T con niveles intermedios y bajos respectivamente,



aunque no de manera estadísticamente significativa (Hull *et al.* 2000). En el presente estudio también se observa que los genotipos A/A y T/A muestran una tendencia a un efecto de riesgo y protección respectivamente, aunque tampoco de manera estadísticamente significativa.

También se ha evaluado la asociación de rs4073 con enfermedades de origen bacteriano, como los estudios de infección por *Clostridium difficile* (CDI) y diarrea por *Clostridium difficile* (CDD). En ambos casos el genotipo A/A presenta un efecto de riesgo, mientras el genotipo T/A muestra un efecto protector par CDD (Garey *et al.* 2010, Jiang *et al.* 2006).

Para confirmar la funcionalidad de este SNP, en un reciente estudio se utilizó la tecnología CRISPR/Cas9 para generar líneas celulares que contengan haplotipos específicos de IL-8 y evaluar los cambios de expresión del gen y la proteína. Al editar los SNPs, rs4073 y rs2227543 y rs2227306 de la línea celular HRK293T, se generó un haplotipo ATC/TTC, que incrementó los niveles de transcripción y traducción de IL-8 de manera diferencial (Benakanakere *et al.* 2017). Estos resultados ponen de manifiesto la participación o papel del alelo A sobre la modulación de la expresión del gen.

La ubicación del SNP rs4073 en la región promotora permite indagar sobre su función biológica. Se sabe que la región promotora de *IL-8* tiene sitios de unión al factor de transcripción NF-kB y a AP-1 que regulan la transcripción del gen (Bezerri *et al.* 2011), acción que había sido previamente evaluada en células infectadas con HTLV-1 en función de la expresión de Tax (Mori *et al.* 1998). Hasta el momento no se ha evaluado si la mayor expresión del gen *IL-8*, mediada por el alelo A del SNP rs4073, se ve exacerbada en células infectadas con HTLV-1. Por otro lado, es importante recordar que la ubicación de este SNP (-251) es *downstream* a los sitios de unión de NF-kB (-94/-81) lo que dificulta una interacción directa entre SNP y el factor de transcripción. Sin embargo, encontrarse en desequilibrio de ligamiento con otros SNPs funcionales en la región promotora podría

generar una acción sinérgica que modifique la transcripción del gen, como ya se ha reportado con la presencia de haplotipos en las regiones promotoras y reguladoras del gen MGMT (Xu *et al.* 2016). Adicionalmente, se ha propuesto como factor de transcripción putativo de *IL-8* a eEF1A1, la cual tendría un sitio de unión en la ubicación del SNP rs4073 (Ahn *et al.* 2011), lo cual resulta interesante pues dicha proteína está envuelta en la producción de citocinas de Th1 (Yan *et al.* 2008).

#### *6.4 Análisis de asociación de haplotipos*

La genotipificación de los SNPs rs2227543 y rs4073 y el hecho de que estos se encuentren en LD (Figura 7) permiten el análisis de haplotipos, con lo cual se evaluó la frecuencia de las distintas combinaciones de alelos entre los grupos PA y PET, así como sus OR para la presencia de la enfermedad.

Tanto el análisis crudo como los análisis ajustados 1 y 2 muestran que ningún haplotipo llega a tener un efecto protector o de riesgo sobre la enfermedad (Tablas 14, 15 y 16), por lo que no se encuentra evidencia de una acción sinérgica entre ambos SNPs. No obstante, hay otros estudios en los que sí se ha reportado la acción de ambos SNPs en haplotipos asociados a enfermedades inflamatorias (Hacking *et al.* 2004). La diferencia de estos resultados puede deberse a que en el estudio de Hacking los haplotipos estaban conformados por un mayor número de SNPs, aumentando el número de posibles combinaciones y la probabilidad de que algunos sean más funcionales que los evaluados en este estudio. Asimismo, es importante recordar que la etiología y desarrollo de la enfermedad es distinta a la de PET, por lo que incluso los factores de riesgo a desarrollar la enfermedad son diferentes.

#### *6.5 Relación entre la carga proviral y las variables evaluadas*

Respecto a la carga proviral, los valores más altos se encuentran en el grupo PET, mientras que los más bajos en PA (Figura 12). Esta diferencia significativa es ampliamente reportada, por lo que el alta carga proviral es uno de los pocos marcadores actuales que se posee para predecir la enfermedad.

A pesar de no ser el objetivo principal de este estudio evaluar la relación entre los genotipos y la carga proviral, el análisis de regresión lineal nos permite ver si hay una relación de esta variable con los SNPs y con las otras variables sean tanto categóricas como numéricas.

Como se esperaba, es el estado de la enfermedad quien presenta una relación fuerte con la carga proviral (coeficiente y OR). Esto se relaciona con el hecho de que la carga proviral es un indicador clínico del desarrollo de la enfermedad aunque no en su totalidad.

Una de las posibles causas que explican la presencia de enfermedad en portadores con alta carga proviral es que un mayor número de células infectadas aumenta las probabilidades de infección no solo mediante la liberación de viriones, sino también por sinapsis viral (Mazurov *et al.* 2010). Ante una mayor presencia de provirus las vías reguladas tanto por Tax, Rex y las proteínas accesorias estarían más activas, siendo algunas vías como la NF- $\kappa$ B la responsable de activar la expresión de los genes responsables de la respuesta inflamatoria que deriva en PET.

En estudios previos, ya se ha demostrado que la carga proviral está asociada con la activación del sistema inmune en portadores asintomáticos. Una alta frecuencia de linfocitos CD4<sup>+</sup> activados expresando CD25<sup>+</sup>CD45RO<sup>+</sup>, HLA-DR<sup>+</sup> y células produciendo FN- $\gamma$  ha sido observada en individuos con una carga proviral mayor al 1% de células infectadas, comparando con portadores asintomáticos (Coutinho *et al.* 2014). A pesar de ello es difícil especificar la influencia de la carga proviral sobre la respuesta inmune

porque varias citoquinas proinflamatorias altamente exacerbadas han sido encontradas mayormente en pacientes con PET pero también en algunos pacientes asintomáticos (Santos *et al.* 2004)

Por otro lado, se postula que la exacerbación de linfocitos T CD4<sup>+</sup> activados por el virus se deba a la poca presencia o acción de los linfocitos T CD8<sup>+</sup>, que son los encargados de destruir a las células infectadas. Esto puede deberse a que la alta carga proviral degenera el reconocimiento por los linfocitos T CD8<sup>+</sup> (Kubota *et al.* 2003) o también a que debido a una mayor cantidad de carga proviral aumenta la probabilidad de que estos linfocitos T CD8<sup>+</sup> también sean infectados con el virus y así este pueda escapar de la respuesta inmune innata para desarrollar la enfermedad.

Uno de los genotipos que parece tener una relación positiva con la carga proviral es T/A del SNP rs4073 (*p-value*: 0.0526). Si bien es interesante cómo este SNP mantiene una tendencia a un efecto protector aun pudiendo presentar valores altos de carga proviral, sería un error relacionarlo a un efecto real debido a la incertidumbre generada por la amplitud del intervalo de confianza (Tabla 11)

#### *6.6 Consideraciones del estudio de asociación genética*

El Perú es un país pluricultural cuya historia reporta migración europea, asiática y africana así como la presencia de comunidades nativas autóctonas. Esto sugiere la presencia de etnias en la población. No obstante, el presente estudio no reporta estratificación poblacional. Visualmente esto se manifiesta como la pérdida de correlación entre los valores eig (Anexo 4 y 5) los cuales han sido obtenidos de un estudio previo donde se utilizaron 37 SNPs como marcadores informativos de ancestría (AIMs) (Talledo *et al.* 2012)

También, en esta investigación se cuenta con un gran número de portadores del virus en la muestra (385) entre asintomáticos y PET, en comparación a otras investigaciones realizadas para la misma enfermedad (Catalan-Soares *et al.* 2009) o enfocados en otras enfermedades inflamatorias (Hacking *et al.* 2004, Hull *et al.* 2000) donde el tamaño de la muestra no supera los 100 individuos. Tomando en cuenta la baja prevalencia de la enfermedad (2% del total de portadores) y la baja frecuencia alélica de algunos SNPs en la población, es necesario contar con un número de individuos que nos permita realizar un análisis de asociación adecuado.

Sin embargo, es necesario considerar que en el presente estudio ambos SNPs no cuentan con un poder estadístico mayor al 40%, lo cual hace que este resultado pueda variar al incrementar el número de portadores asintomáticos pues se sugiere una proporción controles:casos de 3:1. De esta forma, posiblemente, se podría corroborar la tendencia de los genotipos del SNP rs4073, de modo que se tenga evidencia sólida, la acción del genotipo T/T en rs2227543.

Además, la Figura 10 muestra que la frecuencia de individuos mayores de 45 años es considerablemente grande. No obstante, no sabemos si estas personas desarrollarán en algún momento de su vida la enfermedad a pesar de mantener la misma condición genética que la hallada en este estudio. Al no ser este un estudio de cohorte, no tenemos información como para detectar la presencia de estos “falsos asintomáticos” que podrían variar en un futuro los resultados obtenidos en la actualidad.

## 7. CONCLUSIONES

Aunque se observa una tendencia de efecto protector del genotipo T/A del SNP rs4073 y efecto de riesgo del genotipo T/T del SNP rs2227543, no se encuentra evidencia estadísticamente significativa de asociación con el desarrollo de PET. La ausencia de asociación entre los SNPs y la presencia de enfermedad se mantiene aún en presencia de los factores de riesgo.

Las frecuencias alélicas del SNP rs2227543 y SNP rs4073 son  $MAF=0.17$  y  $MAF=0.24$ , respectivamente. Además, los genotipos al estar en equilibrio de Hardy-Weinberg, tienen frecuencias genotípicas de  $p^2$ ,  $2pq$  y  $q^2$  donde  $p=MAF$ .

No se encuentra evidencia de que las combinaciones haplotípicas entre el SNP rs2227543 y rs4073 se asocien con el desarrollo de la enfermedad en ausencia o presencia de factores de riesgo.

No se encuentra evidencia de asociación por parte de ningún genotipo de cualquiera de los dos SNPs con la carga proviral.

## 8. RECOMENDACIONES

Se sugiere ampliar el estudio aumentando el número de individuos asintomáticos para llegar a una proporción 3:1 entre controles y casos, como se aconseja para los estudios de asociación de diseño caso control.

Evaluar la acción del SNP rs4073 junto a otros polimorfismos presentes en la región promotora de *IL-8*. Una acción sinérgica podría aumentar el efecto de rs4073 como de otros SNPs y generar haplotipos que puedan servir como un marcador genético para el diagnóstico temprano de paraparesia espástica tropical.

Evaluar la interacción y funcionalidad de eEF1A1 y el SNP rs4073, para corroborar su acción como posible factor de transcripción de *IL-8*.

## 9. BIBLIOGRAFÍA

- AHN MH, PARK BL, LEE SH, PARK SW, ET AL. A promoter SNP rs4073T>A in the common allele of the interleukin 8 gene is associated with the development of idiopathic pulmonary fibrosis via the IL-8 protein enhancing mode. *Respiratory Research*. 2011, 12:73
- ARAYA N, SATO T, YAGISHITA N, ANDO H, ET AL. Human T-lymphotropic virus type 1 (HTLV-1) and regulatory T cells in HTLV-1-associated neuroinflammatory disease. *Viruses*. 2011;3(9):1532-48.
- BANGHAM CRM, ARAUJO A, YAMANO Y y TAYLOR GP. HTLV-1-associated myelopathy/tropical spastic paraparesis. *Nature Reviews Disease Primers*. 2015 6;15023.
- BENAKANAKERE MR, FINOTI LS, TANAKA U, GRANT GR, ET AL. Investigation of the functional role of human Interleukin-8 gene haplotypes by CRISPR/Cas9 mediated genome editing. *Sci. Rep.*2016. 6, 31180.
- BEZZERRI V, BORGATTI M, FINOTTI A, TAMANINI A, ET AL. Mapping the Transcriptional Machinery of the IL-8 Gene in Human Bronchial Epithelial Cells. *J Immunol*. 2011, 187 (11) 6069-6081.
- BLATTNER WA. Epidemiology of HTLV-I and associated diseases. In: Blattner WA, ed. Human retrovirology: HTLV. *New York: Raven Press*. 1990.
- BORILOVA LINHARTOVA P, VOKURKA J, POSKEROVA H, FASSMANN A, ET AL. Haplotype Analysis of Interleukin-8 Gene Polymorphisms in Chronic and Aggressive Periodontitis. *Mediators of Inflammation*. 2013;2013:8.
- BRAT DJ, BELLAIL AC y VAN MEIR EG. The role of interleukin-8 and its receptors in gliomagenesis and tumoral angiogenesis. *Neuro Oncol*. 2005; 7(2): 122–133.
- BROGLIO P. Neuroimmunology Research Focus. *Nova biomedics*; 2007. 43p.
- CASSAR O, EINSIEDEL L, AFONSO P y GESSAIN A. Human T-Cell Lymphotropic Virus Type 1 Subtype C Molecular Variants among Indigenous Australians: New Insights into the Molecular Epidemiology of HTLV-1 in Australo-Melanesia. *PLoS Negl Trop Dis*. 2013; 7(9): e2418.
- CATALAN-SOARES BC, CARNEIRO-PROIETTI AB, DA FONSECA FG, CORREA-OLIVEIRA R, ET AL. Interdisciplinary HTLV Research Group (GIPH), Proietti FA. HLA



class I alleles in HTLV-1-associated myelopathy and asymptomatic carriers from the Brazilian cohort GIPH. *Med Microbiol Immunol*. 2009;198(1):1-3.

- COLLINS y KE. Primer1: Primer Design Web Service for Tetra-Primer ARMS-PCR. *The Open Bioinformatics Journal*. 2012 6, 55-58.
- CONSEJO NACIONAL DE SALUD DEL PERÚ. Lineamientos y medidas de reforma del sector salud. *Ministerio de salud*. 2013.
- COLER-REILLY A, YAGISHITA N, SUZUKI H, SATO T, ET AL. Nation-wide epidemiological study of Japanese patients with rare viral myelopathy using novel registration system (HAM-net). *Orphanet J Rare Dis*. 2016; 11: 69.
- COOK LB, ROWAN AG, MELAMED A, TAYLOR GP, ET. HTLV-1-infected T cells contain a single integrated provirus in natural infection. *Blood*. 2012;120(17):3488-90.
- COOK LB, ELEMANS M, ROWAN AG y ASQUITH B. HTLV-1: Persistence and pathogenesis. *Virology*. 2013; 435. 131–140.
- COUTINHO R, RIOS G, KORNGOLD A, NILLA V, ET AL. Human T lymphotropic virus type 1 (HTLV-1) proviral load induces activation of T-lymphocytes in asymptomatic carriers. *BMC Infect Dis*. 2014; 14: 453.
- DE CASTRO-COSTA, ARAÚJO A, BARRETO M, TAKAYANAQUI O, ET AL. Proposal for diagnostic criteria of tropical spastic paraparesis/HTLV-I-associated myelopathy (TSP/HAM). *AIDS Research and Human Retroviruses*. 2006, 22(10): 931-935.
- DE LAS CASAS C, GOTUZZO E , DEZA L, CABRERA J, ET AL. Características epidemiológicas de los pacientes con paraparesia espástica tropical (PET) asociada a infección por HTLV-1 en Lima, Perú. *Rev Med Hered*. 1996; 7: 68-74.
- DE FREITAS CARNEIRO AB. *Cadernos Hemominas HTLV*. Fundación Hemominas. 5ta edición. Belo Horizonte 2010. 11 p. ISBN. 978-85-60055-03-6
- DE THE G y BOMFORD R. An HTLV-I vaccine: why, how, for whom? *AIDS Res. Hum. Retroviruses*. 1993,9, 381–386.
- ECHEVERRI y MOCKUS. Factor Nuclear kB (NF-κB): Signalosoma y su importancia en enfermedades inflamatorias y cáncer. *rev.fac.med* 2008. vol.56 no.2
- FINDLEY, T. W. sed3: Variable transformation and evaluation. Stata Technical Bulletin 2: 15. *Reprinted in Stata Technical Bulletin Reprints* 1990. vol. 1, pp. 85–86.

- FLORES-ALFARO E, BURGUETE-GARCÍA AI y SALAZARMARTÍNEZ E. Diseños de investigación en epidemiología genética. *Rev Panam Salud Publica*. 2012;31(1):88–94.
- FURTADO MDOS S, ANDRADE RG, ROMANELLI LC, RIBEIRO MA, ET AL. Monitoring the HTLV-1 proviral load in the peripheral blood of asymptomatic carriers and patients with HTLV-associated myelopathy/tropical spastic paraparesis from a Brazilian cohort: ROC curve analysis to establish the threshold for risk disease. *J Med Virol*. 2012 Apr;84(4):664-71.
- GADELHA SR, ALCÂNTARA LCJ, COSTA GC, ACOSTA AX, ET AL. Correlation Between Polymorphisms at Interleukin-6 But Not at Interleukin-10 Promoter and the Risk of Human T Lymphotropic Virus Type I-Associated Myelopathy/Tropical Spastic Paraparesis in Brazilian Individuals. *Journal of Medical Virology*. 2008 80:2141–2146.
- GAREY KW, JIANG Z-D, GHANTOJI S, TAM VH, ARORA V, ET AL. A Common Polymorphism in the Interleukin-8 Gene Promoter Is Associated with an Increased Risk for Recurrent Clostridium difficile Infection. *Clinical Infectious Diseases*. 2010 Dec 15;51(12):1406–10.
- GESSAIN A y CASSAR O. Epidemiological Aspects and World Distribution of HTLV-1 Infection. *Front Microbiol*. 2012;3.
- GONCALVES DU, PROIETTI FA, BARBOSA-STANCIOLI EF, MARTINS ML, ET AL. HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP) inflammatory network. *Inflamm Allergy Drug Targets*. 2008 Jun;7(2):98-107.
- GONZÁLEZ JR, ARMENGOL L, GUINÓ E, SOLÉ X y MORENO V (2014). SNPAssoc: SNPs-based whole genome association studies. R package version 1.9-2. <https://CRAN.R-project.org/package=SNPAssoc>
- GOTUZZO E, GONZÁLEZ E, VERDONCK K, MAYER E, ET AL. Veinte años de investigación sobre HTLV-1 y sus complicaciones médicas en el Perú: Perspectivas generales. *Acta méd. peruana* [online]. 2010; 27(3):196-203.
- GOTUZZO E, VERDONCK K, GONZALEZ E y CABADA M. Virus linfotrópico humano de células T tipo 1 (HTLV-1): Una infección endémica en el Perú. *Rev. Perú. med. exp. salud pública* [online]. 2004; 21(4):253-260.
- GRIMM MC, ELSBURY S K, PAVLI P y DOE W F. Interleukin 8: cells of origin in inflammatory bowel disease. *Gut*. 1996 Jan; 38(1): 90–98.
- GROSS y LIGGES (2015). nortest: Tests for Normality. R package version 1.0-4. <https://CRAN.R-project.org/package=nortest>

- HACKING D, KNIGHT JC, ROCKETT K, BROWN H, ET AL. Increased in vivo transcription of an IL-8 haplotype associated with respiratory syncytial virus disease-susceptibility. *Genes Immun.* 2004 Apr 15;5(4):274–82
- HLELA C, SHEPPERD S, KHUMALO NP, TAYLOR GP. The prevalence of human T-cell lymphotropic virus type 1 in the general population is unknown. *AIDS* 2009, Rev. 11, 205–214.
- HUANG Y, YU H, CAO Q, ET AL. The Association of Chemokine Gene Polymorphisms with VKH and Behcet's Disease in a Chinese Han Population. *BioMed Research International.* 2017;2017:1274960.
- HULL J, THOMSON A y KWIATKOWSKI D. Association of respiratory syncytial virus bronchiolitis with the interleukin 8 gene region in UK families. *Thorax.* 2000 Dec;55(12):1023–7.
- IKRAM ULLAH, TOOBA IFTIKHAR, MUHAMMAD AYUB JADOON, ET AL. Heterogeneous geographic distribution of human Tcell lymphotropic virus I (HTLV-I) and epidemiological, genetical, pathophysiological, therapeutic knowledge about HTLV. *Journal of Entomology and Zoology Studies* 2017; 5(2): 871-878.
- IZUMO, S. Neuropathology of HTLV-1-associated myelopathy (HAM/TSP). *Neuropathology*, 2010. 30: 480–485.
- JEFFERY KJ, USUKU K, HALL SE, MATSUMOTO W, ET AL. HLA alleles determine human T-lymphotropic virus-I (HTLV-I) proviral load and the risk of HTLV-I-associated myelopathy. *Proc Natl Acad Sci U S A.* 1999 Mar 30;96(7):3848-53.
- JIANG ZD, DUPONT HL, GAREY K, ET AL. A common polymorphism in the interleukin 8 gene promoter is associated with Clostridium difficile diarrhea. *Am J Gastroenterol* 2006; 101:1112–1116.
- KAMOI K, MOCHIZUKI M. HTLV-1 uveitis. *Frontiers in Microbiology.* 2012;3:270.
- KASLOW RA, OSTROW DG, DETELS R, PHAIR JP, ET AL. The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. *Am J Epidemiol.* 1987;126(2):310–8.
- KENDALL E, GONZÁLEZ E, ESPINOZA I, TIPISMANA M, ET AL. Early neurologic abnormalities associated with human T-cell lymphotropic virus type 1 infection in a cohort of Peruvian children. *The journal of pediatrics* 2009; 155(5):700-6.

- KHOURY MJ y BEATY TH. Aplicaciones del método de casos y controles en epidemiología genética. *Bol Of Sanit Panam*.1996;121(5):421–40.
- KUBOTA R1, FURUKAWA Y, IZUMO S, USUKU K, ET AL. Degenerate specificity of HTLV-1-specific CD8+ T cells during viral replication in patients with HTLV-1-associated myelopathy (HAM/TSP). *Blood*. 2003 Apr 15;101(8):3074-81.
- KUNSCH C y ROSEN C. NF-kappa B subunit-specific regulation of the interleukin-8 promoter. *Molecular and Cell Biology* 1993; 13(10):6137-46.
- LEWIS C y KNIGHT J. Introduction to genetic Association Studies. *Cold Spring Harbor protocols*. 2009.
- LIU HH, HU Y, ZHENG M, SUHOSKI MM, ENGLEMAN EG, ET AL. Cd14 SNPs regulate the innate immune response. *Mol Immunol*. 2012 Jun;51(2):112-27.
- MAHIEUX y GESSAIN. HTLV-3/STLV-3 and HTLV-4 Viruses: Discovery, Epidemiology, Serology and Molecular Aspects. *Viruses*. 2011 Jul; 3(7): 1074–1090.
- MALONEY EM, CLEGHORN FR, MORGAN OS, RODGERS-JOHNSON P, ET AL. Incidence of HTLV-I-associated myelopathy/tropical spastic paraparesis (HAM/TSP) in Jamaica and Trinidad. *J Acquir Immune Defic Syndr Hum Retrovirol*. 1998 Feb 1;17(2):167-70.
- MARTIN F, FEDINA A, YOUSHYA S, TAYLOR GP. A 15-year prospective longitudinal study of disease progression in patients with HTLV-1 associated myelopathy in the UK. *J Neurol Neurosurg Psychiatry*. 2010 Dec;81(12):1336-40.
- MATHESON MC, ELLIS JA, RAVEN J, WALTERS EH, ET AL. Association of IL8, CXCR2 and TNF-[alpha] polymorphisms and airway disease. *J Hum Genet*. 2006 Jan 21;51(3):196–203.
- MATSUOKA y JEANG. Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation. *Nat Rev Cancer*. 2007 Apr;7(4):270-80.
- MATSUOKA y JEANG. Human T-cell leukemia virus type 1 (HTLV-1) and leukemic transformation: viral infectivity, Tax, HBZ and therapy. *Oncogene* (2011) 30, 1379–1389.
- MATSUOKA M y YASUNAGA J. Human T-cell leukemia virus type 1: replication, proliferation and propagation by Tax and HTLV-1 bZIP factor. *Curr Opin Virol*. 2013 Sep 21. 3(6), 684-691.

- MAZUROV D, ILINSKAYA A, HEIDECKER G, LLOYD P, ET AL.. Quantitative comparison of HTLV-1 and HIV-1 cell-to-cell infection with new replication dependent vectors. *PLoS Pathog.* 2010 Feb 26;6(2):e1000788.
- MILLAR D, HORAN M, CHUZHANOVA N, COOPER D. Characterisation of a functional intronic polymorphism in the human growth hormone (GHI) gene. *Hum Genomics.* 2010; 4(5): 289–301.
- MINSA. Norma técnica de salud para el manejo de las ITS. Dirección General de Salud de las Personas. Estrategia Sanitaria Nacional Prevención y Control de Infecciones de Transmisión Sexual y VIH-SIDA-Lima. *Ministerio de Salud*; 2006.
- MORI N, MUKAIDA N, BALLARD D, MATSUSHIMA K, ET AL. Human T-Cell Leukemia Virus Type I Tax Transactivates Human Interleukin 8 Gene through Acting Concurrently on AP-1 and Nuclear Factor-KB-like Sites. *Cancer research* 1998; 58(17):3993-4000.
- NAKANO K y WATANABE T. HTLV-1 Rex Tunes the Cellular Environment Favorable for Viral Replication. *Viruses.* 2016 Mar; 8(3): 58.
- OH U, MCCORMICK MJ, DATTA D, ET AL. Inhibition of immune activation by a novel nuclear factor-kappa B inhibitor in HTLV-I-associated neurologic disease. *Blood.* 2011;117(12):3363-3369.
- ORGANIZACIÓN MUNDIAL DE LA SALUD. Políticas de la OMS en materia de género 2002.
- ORTEGA V y MEYERS D. Pharmacogenetics: Implications of Race and Ethnicity on Defining Genetic Profiles for Personalized Medicine. *J Allergy Clin Immunol.* 2014 January; 133(1): 16–26.
- OSAME M. 1990. Review of WHO Kagoshima meeting and diagnostic guidelines for HAM/TSP. In: Blattner WA, editor. Human retrovirology: HTLV. New York: Raven Press, pp 191–197.
- PATTERSON N, PRICE AL, REICH D. Population Structure and Eigenanalysis. *PLoS Genet.* 2006 Dec;2(12):e190.
- PELOPONESE J-M JR., YEUNG ML, JEANG K-T. Modulation of Nuclear Factor-Kb by Human T Cell Leukemia Virus Type 1 Tax Protein. *Immunologic Research* 2006;34/1:1–12.
- POIESZ BJ, RUSCETTI FW, GAZDAR AF, BUNN PA, ET AL. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with

cutaneous T-cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America*. 1980;77(12):7415-7419.

- QAZI BS, TANG K, QAZI A. Recent Advances in Underlying Pathologies Provide Insight into Interleukin-8 Expression-Mediated Inflammation and Angiogenesis. *International Journal of Inflammation*. 2011;2011:908468.
- QUARESMA J, YOSHIKAWA G, KOYAMA R, DIAS G, ET AL HTLV-1, Immune Response and Autoimmunity. *Viruses*. 2016 Jan; 8(1):5.
- RAMÍREZ-BELLO J, VARGAS-ALARCÓN G, TOVILLA-ZÁRATE C, FRAGOSO JM. Polimorfismos de un solo nucleótido (SNP): implicaciones funcionales de los SNP reguladores (rsNP) y de los SNP-ARN estructurales (srNP) en enfermedades complejas. *Gaceta médica de México* 2013; 149(2):220-8.
- REVELLE, W. (2017) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, <https://CRAN.R-project.org/package=psych> Version = 1.7.5.
- RINGNÉR M. What is principal component analysis? *Nature Biotechnology* 2008. 26, 303 – 304.
- ROSSI D y ZLOTNIK A. The biology of chemokines and their receptors. *Annu Rev Immunol*. 2000;18:217-42.
- SABOURI AH, SAITO M, LLOYD AL, VINE AM, ET AL. Polymorphism in the Interleukin-10 Promoter Affects Both Provirus Load and the Risk of Human T Lymphotropic Virus Type I–Associated Myelopathy/Tropical Spastic Paraparesis. *The Journal of Infectious Diseases* 2004; 190:1279–85.
- SAITO M, MATSUZAKI T, SATOU Y, ET AL. In vivo expression of the HBZ gene of HTLV-1 correlates with proviral load, inflammatory markers and disease severity in HTLV-1 associated myelopathy/tropical spastic paraparesis (HAM/TSP). *Retrovirology*. 2009;6:19.
- SANTOS SB, PORTO AF, MUNIZ AL, DE JESUS AR, ET AL. Exacerbated inflammatory cellular immune response characteristics of HAM/TSP is observed in a large proportion of HTLV-I asymptomatic carriers. *BMC Infect Dis*. 2004 Mar 2;4:7.
- SAVAGE S, HOU L, LISSOWSKA J, CHOW WH, ET AL. Interleukin-8 Polymorphisms Are Not Associated with Gastric Cancer Risk in a Polish Population. *Cancer Epidemiol Biomarkers Prev* 2006;15(3).
- SILVERMAN EK. Haplotype Thinking in Lung Disease. *Proc Am Thorac Soc*. 2007 Jan; 4(1): 4–8.

- SKOL AD, SCOTT LJ, ABECASIS GR y BOEHNKE M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* 2006;38:209–213.
- SLATKIN M. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nature reviews Genetics.* 2008;9(6):477-485.
- SOLÉ X, GUINÓ E, VALLS J, INIESTA R, ET AL. SNPStats: a web tool for the analysis of association studies. *Bioinformatics* 2006; 22(15):1928-9.
- SUH JS, HAHN WH y CHO BS. Polymorphisms of CXCL8 and Its Receptor CXCR2 Contribute to the Development and Progression of Childhood IgA Nephropathy. *Journal of Interferon & Cytokine Research.* March 2011, 31(3): 309-315.
- SZKLO M. Population-based Cohort Studies. *Epidemiol Rev.* 1998;20(1):81–90.
- TALLEDO M. “Determinación de los perfiles de expresión de mRNA de la vía de señalización NF-kB en pacientes con paraparesia espástica tropical y pacientes asintomáticos HTLV-1 positivos”. Asesor: Dr. Jorge Arévalo Zelada. Tesis de Grado de Magister en Bioquímica y Biología Molecular. UPCH, Escuela de postgrado Víctor Alzamora Castro, Lima, 2013.
- TALLEDO M, LÓPEZ G, HUYGHE J, VERDONCK K, ET AL. Evaluation of Host Genetic and Viral Factors as Surrogate Markers for HTLV-1-Associated Myelopathy/Tropical Spastic Paraparesis in Peruvian HTLV-1-Infected Patients. *Journal of Medical Virology.* 2010. 82:460–466
- UNICEF Perú. Lactancia materna, una Fuente de vida. 2006. [https://www.unicef.org/peru/\\_files/notas\\_prensa/carpetasinformativas/lactancia\\_materna.pdf](https://www.unicef.org/peru/_files/notas_prensa/carpetasinformativas/lactancia_materna.pdf).
- URETA-VIDAL A, ANGELIN-DUCLOS C, TORTEVOYE P, MURPHY E, ET AL. Mother-to-child transmission of human T-cell-leukemia/lymphoma virus type I: Implication of high antiviral antibody titer and high proviral load in carrier mothers. *Int. J. Cancer.* 1999. 82: 832–836.
- VALERI VW, HRYNIEWICZ A, ANDRESEN V, ET AL. Requirement of the human T-cell leukemia virus p12 and p30 products for infectivity of human dendritic cells and macaques but not rabbits. *Blood.* 2010;116(19):3809-3817.
- VERDONCK K, GONZÁLEZ E, VAN DOOREN S, VANDAMME AM, ET AL. Human T-lymphotropic virus 1: recent knowledge about an ancient infection. *Lancet Infect Dis.* 2007 Apr;7(4):266-81.

- VILLAVERDE J, ROMANÍ F, MONTANO S, ZUNT J. Transmisión Vertical De Htlv-1 En El Perú. *Rev Peru Med Exp Salud Publica*. 2011 March; 28(1): 101–108.
- WANG C, ZHAN X, BRAGG-GRESHAM J, MIN H, ET AL. Ancestry Estimation and Control of Population Stratification for Sequence-based Association Studies. *Nat Genet*. 2014 Apr; 46(4): 409–415.
- XU M., CROSS C.E, SPEIDEL J.T. ET AL. MGMT DNA repair gene promoter/enhancer haplotypes alter transcription factor binding and gene expression. *Cellular Oncology*. 2016; 39 (5):435-447.
- YAN G, YOU B, CHEN S-P, LIAO JK, ET AL. TNF-alpha Downregulates Endothelial Nitric Oxide Synthase mRNA Stability via Translation Elongation Factor 1-alpha 1. *Circulation research*. 2008;103(6):591-597.
- YATES A, AKANNI W, RIDWAN AMODE M, BARRELL D, ET AL. Ensembl 2016. *Nucleic Acids Res*. 2016 44 Database issue:D710-6. Disponible en Web: <http://www.ensembl.org/index.html>.
- YE S, DHILLON S, KE X, COLLINS A, DAY I. An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic acids research* 2001; 29(17):E88-8.
- ZAMUDIO R, PEREIRA L, ROCHA C, BERG D, MUNIZ-QUEIROZ T EL AL. Population, Epidemiological, and Functional Genetics of Gastric Cancer Candidate Genes in Peruvians with Predominant Amerindian Ancestry. *Dig Dis Sci* (2016) 61:107–116.
- ZHANG X, LIN S, YANG Y, RONG L, ET AL. Interleukin-2 and Interleukin-8 Gene Polymorphisms and Acquired Aplastic Anemia Risk in a Chinese Population. *Cell Physiol Biochem* 2017;41:1199-1207.
- ZHANG y SHARMA. Human T-Cell Lymphotropic Virus Type-I tax Gene Induces Interleukin-8 Secretion by Autocrine Mechanism and Has No Effect on Interleukin-16 in Transfected Jurkat Cells. *Biochemical and Biophysical Research Communications* 2000. 273, 865–869.
- ZIEGLER A. y KÖNIG IR. Celebrating the 30th anniversary of genetic epidemiology: how to define our scope? *Genet Epidemiol*. 2014 Jul;38(5):379-80.

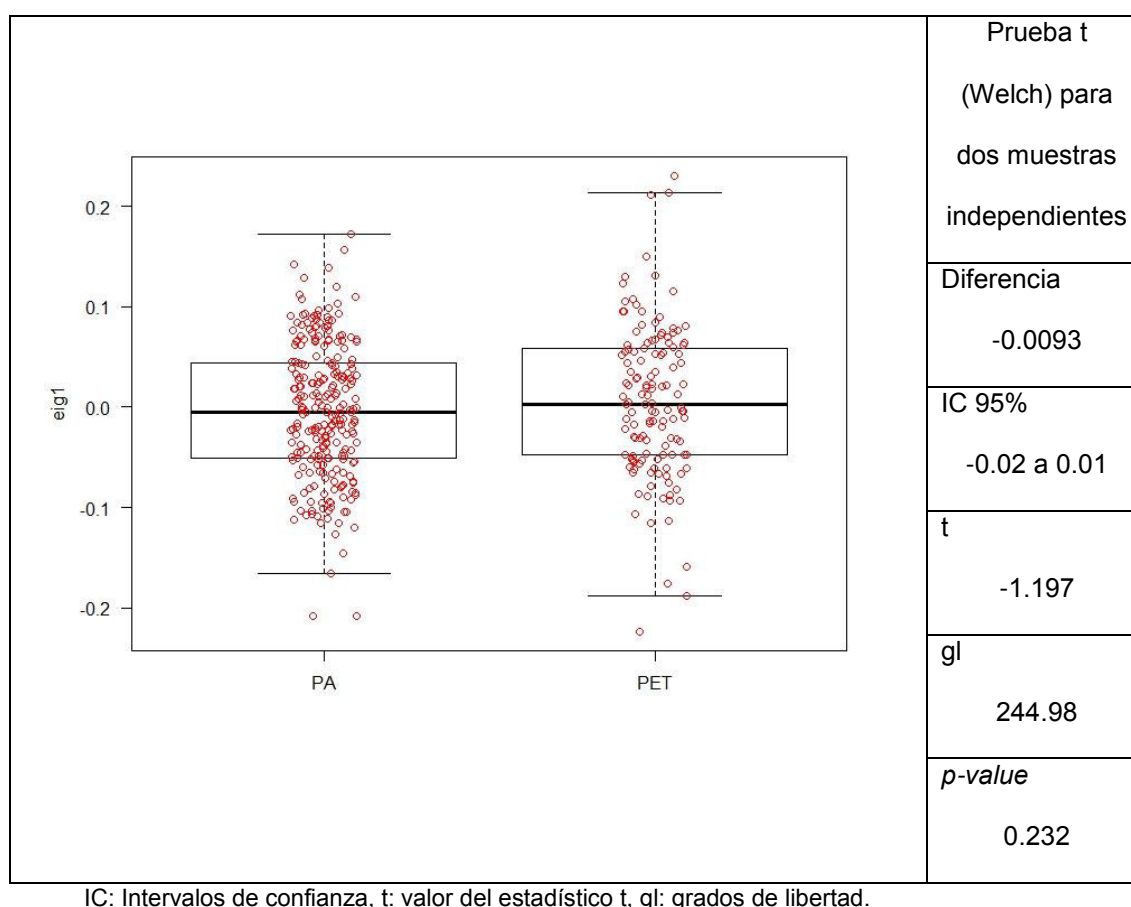


## 10. ANEXOS

## 10.1 ANEXO 1: Descripción de componente principal eig1

**Tabla 18.** Descripción de la variable eig1

eig1	PA	PET
Mínimo	-0.21	-0.22
Máximo	0.17	0.23
Media	-0.0038	0.0055
Mediana	-0.0047	0.0030
Desviación estándar	0.067	0.075
Prueba de normalidad ( <i>p-value</i> )*	0.0719	0.719
*Lilliefors (Kolmogorov-Smirnov)		



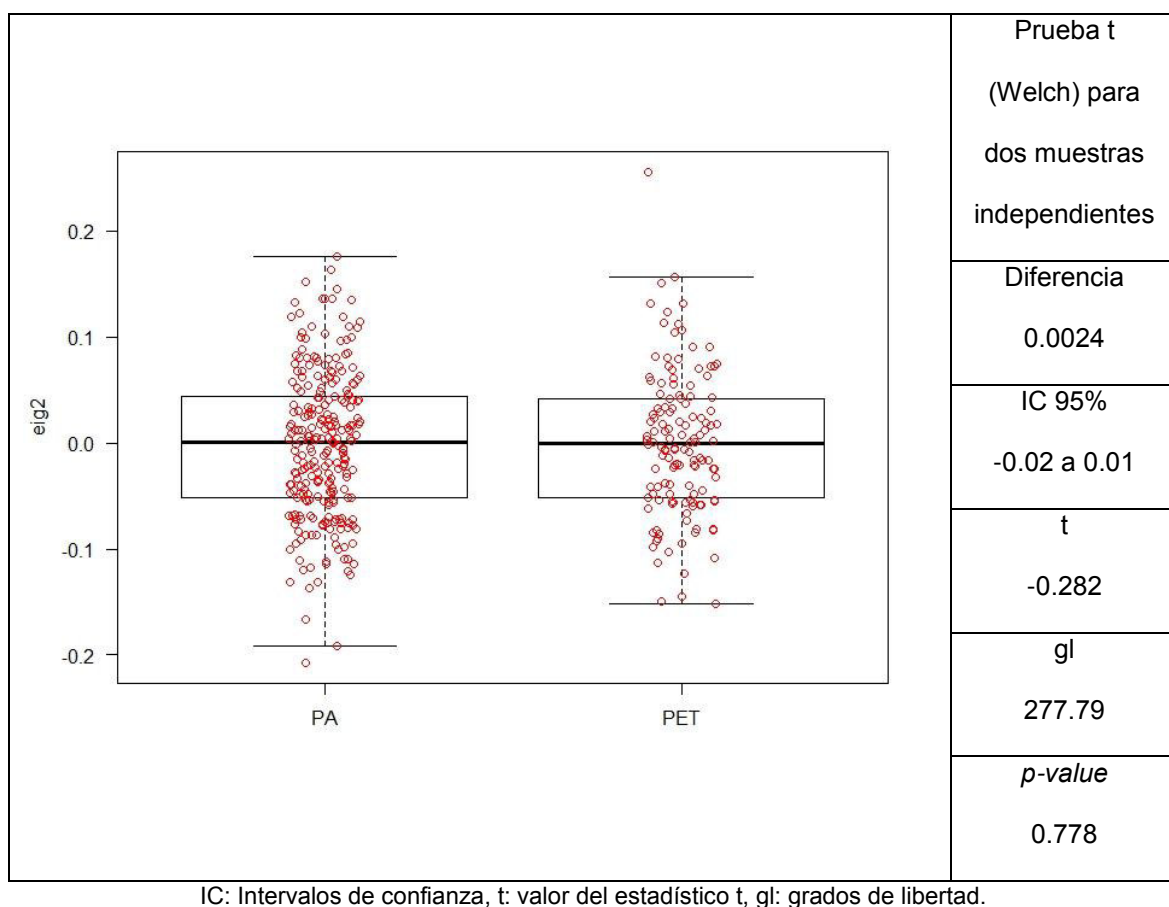
**Figura 19.** Distribución de la variable eig1 entre los grupos PA y PET.

No hay diferencias estadísticamente significativas en la distribución de eig1 entre los grupos PA y PET.

## 10.2 ANEXO 2: Descripción de componente principal eig2

**Tabla 19.** Descripción de la variable eig2

eig2	PA	PET
Mínimo	-0.21	-0.15
Máximo	0.18	0.26
Media	-0.0020	-0.0004
Mediana	0.0007	-0.0005
Desviación estándar	0.069	0.067
Prueba de normalidad ( <i>p-value</i> )*	0.361	0.708
*Lilliefors (Kolmogorov-Smirnov)		



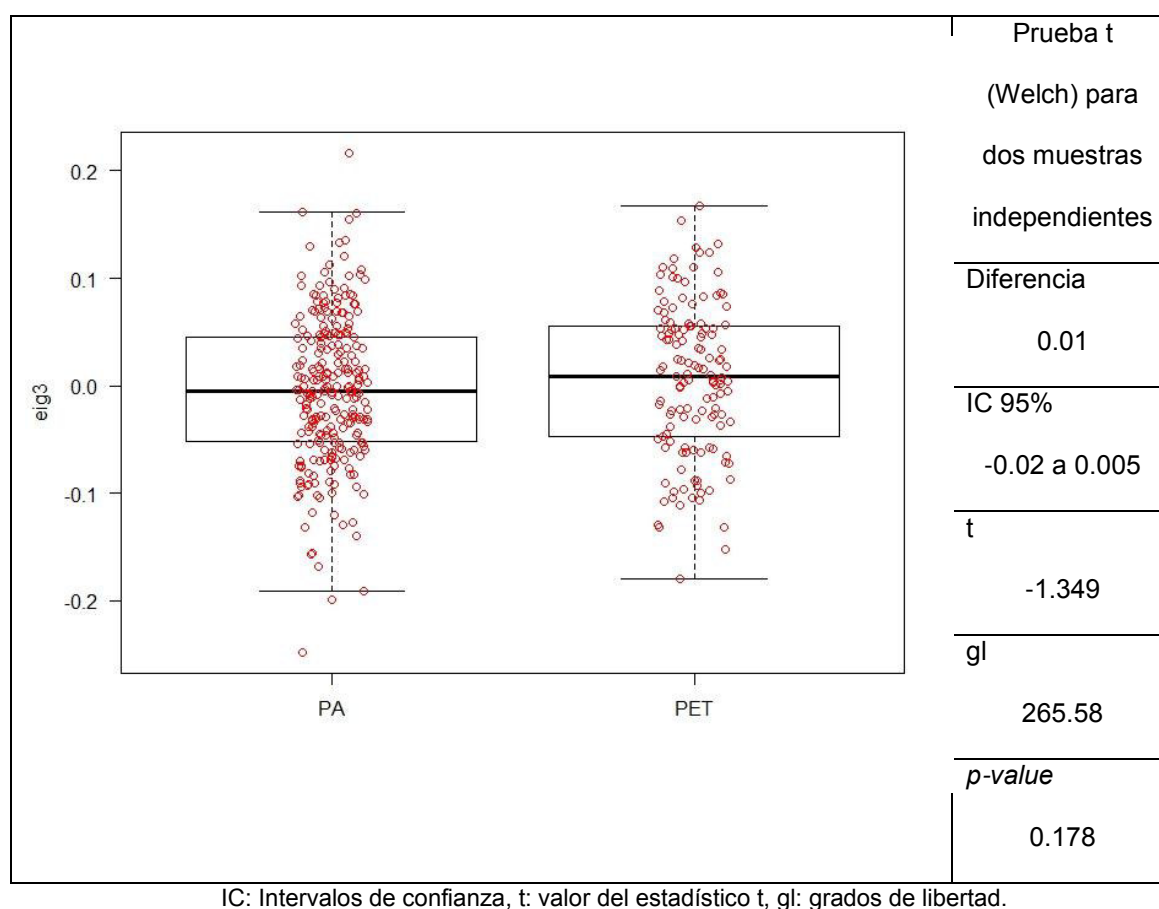
**Figura 20.** Distribución de la variable eig2 entre los grupos PA y PET.

No hay diferencias estadísticamente significativas en la distribución de eig2 entre los grupos PA y PET.

### 10.3 ANEXO 3: Descripción del componente principal eig3

**Tabla 20.** Descripción de la variable eig3

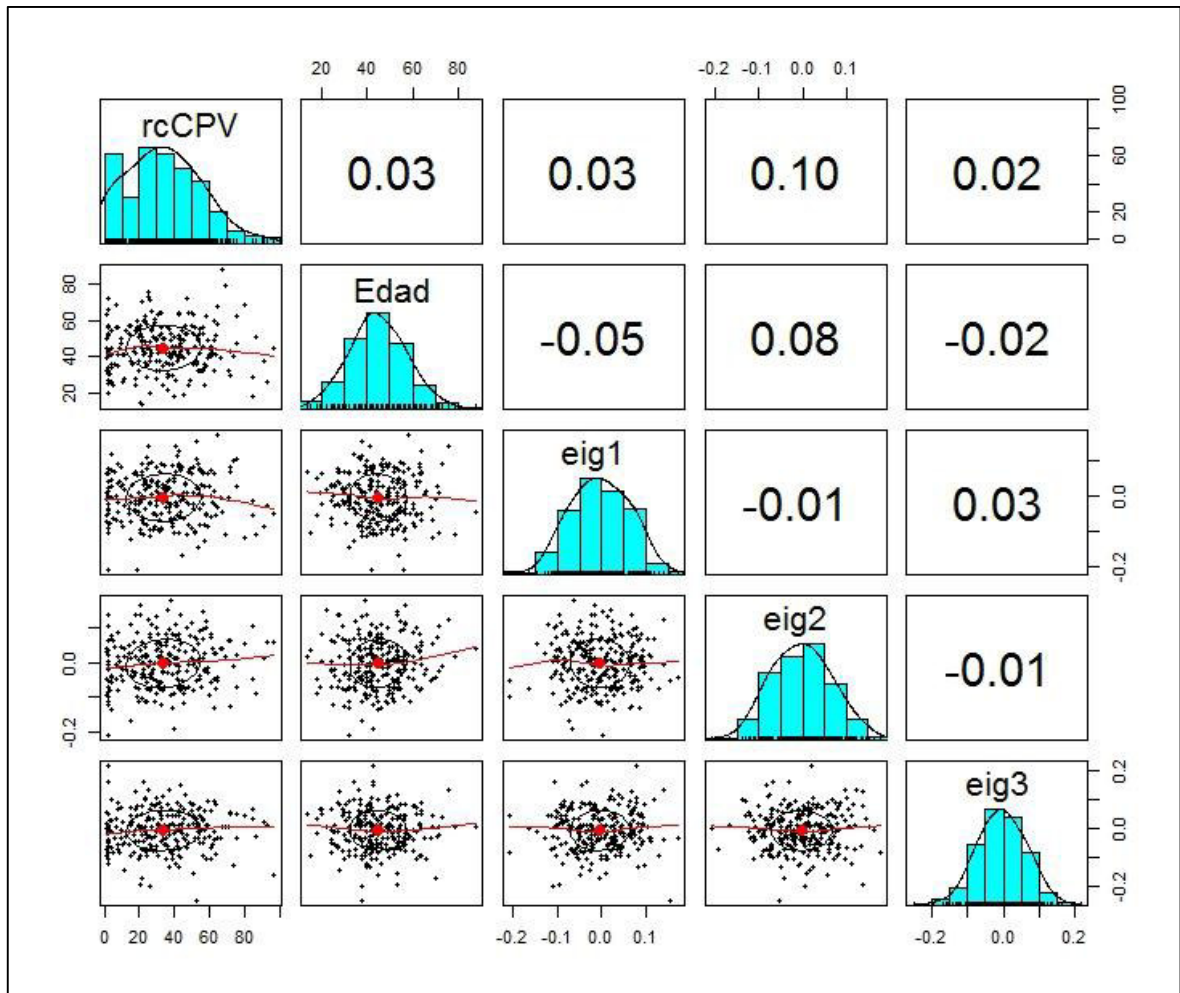
eig3	PA	PET
Mínimo	-0.25	-0.18
Máximo	0.22	0.17
Media	-0.0043	0.0059
Mediana	-0.0046	0.0090
Desviación estándar	0.069	0.071
Prueba de normalidad ( <i>p-value</i> )*	0.754	0.328
*Lilliefors (Kolmogorov-Smirnov)		



**Figura 21.** Distribución de la variable eig3 entre los grupos PA y PET.

No hay diferencias estadísticamente significativas en la distribución de eig3 entre los grupos PA y PET.

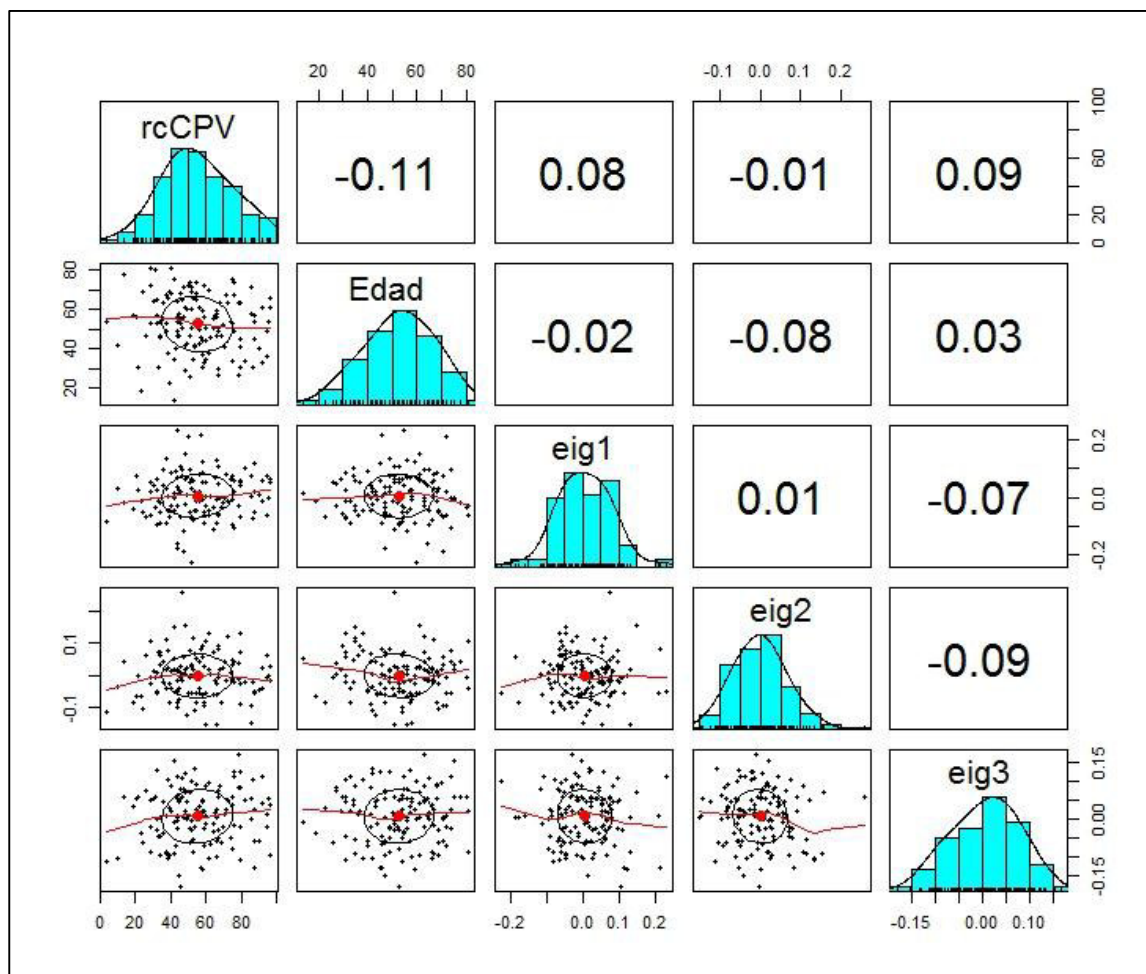
#### 10.4 ANEXO 4: Correlación de variables numéricas en el grupo PA



**Figura 22.** Diagrama de correlación entre las variables numéricas en el grupo PA.

Los diagramas de dispersión se acompañan con líneas suavizadas de tendencia así como elipses de correlación. En la diagonal se indican los histogramas con la curva de densidad. Los coeficientes de correlación de Pearson se muestran en la sección diagonal superior. No hay correlación entre ninguna de las variables por lo que no son dependientes entre sí. Gráfico elaborado con el paquete “psych” en R (Revelle 2017).

## 10.5 ANEXO 5: Correlación de variables numéricas en el grupo PET



**Figura 23.** Diagrama de correlación entre las variables numéricas en el grupo PET.

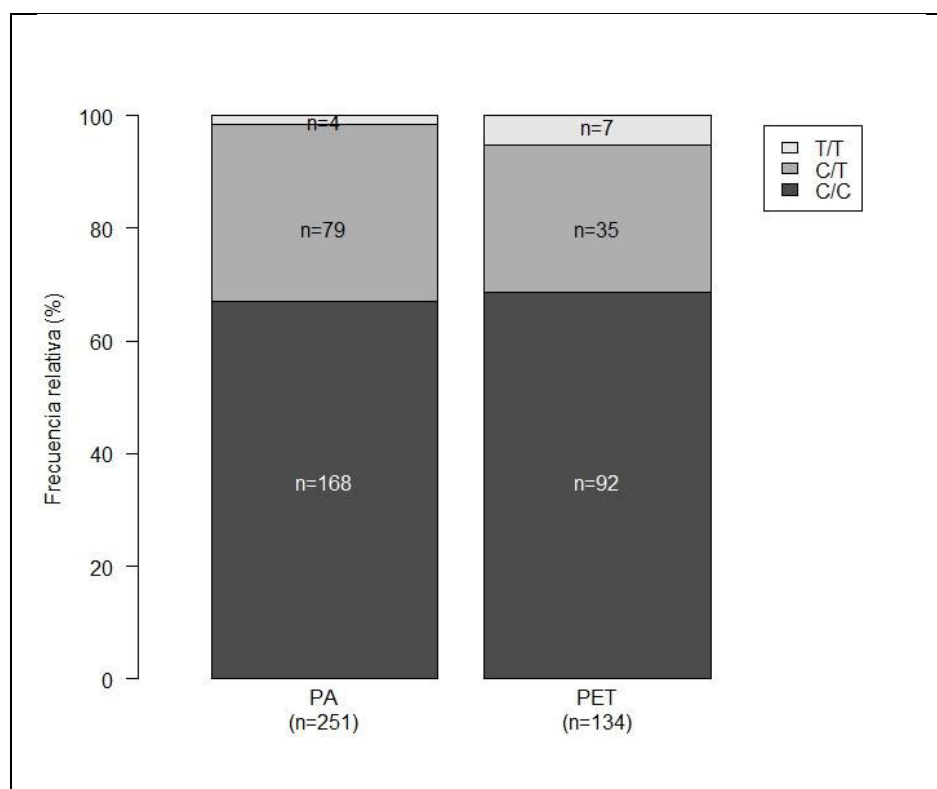
Los diagramas de dispersión se acompañan con líneas suavizadas de tendencia así como elipses de correlación. En la diagonal se indican los histogramas con la curva de densidad. Los coeficientes de correlación de Pearson se muestran en la sección diagonal superior. No hay correlación entre ninguna de las variables por lo que no son dependientes entre sí. Gráfico elaborado con el paquete “psych” en R (Revelle 2017).

## 10.6 ANEXO 6: Frecuencias alélicas y genotípicas del SNP rs2227543

**Tabla 21.** Frecuencias alélicas y genotípicas del SNP rs2227543

Genotipos	PA			PET		
	n (%)	Frecuencias alélicas		n (%)	Frecuencias alélicas	
		T	C		T	C
C/C	168 (67)	0.17	0.83	92 (69)	0.18	0.82
C/T	79 (31)			35 (26)		
T/T	4 (2)			7 (5)		
X <sup>2</sup>	2.431			2.124		
HWE*	0.18			0.15		

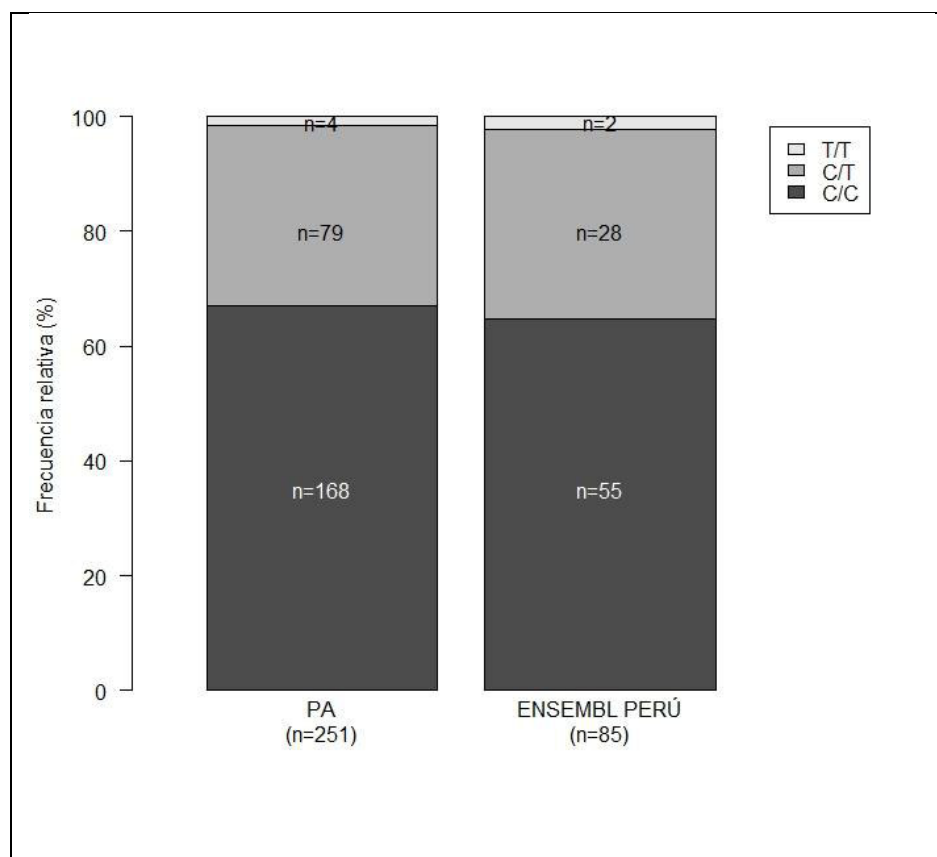
X<sup>2</sup>: Valor de Chi-cuadrado, EHW: Equilibrio de Hardy-Weinberg.



**Figura 24.** Distribución de las frecuencias genotípicas entre los grupos PA y PET del SNP rs2227543.

No hay diferencias estadísticamente significativas (Prueba Chi-cuadrado de Pearson. *p*-value: 0.0883).

10.7 ANEXO 7: Comparación de las frecuencias genotípicas del grupo PA con las frecuencias reportadas por el proyecto Ensembl del SNP rs2227543



**Figura 25.** Comparación de las frecuencias genotípicas del grupo PA con las del proyecto Ensembl para el SNP rs2227543.

No hay diferencias estadísticamente significativas (Prueba de Fisher, *p-value*: 0.758).

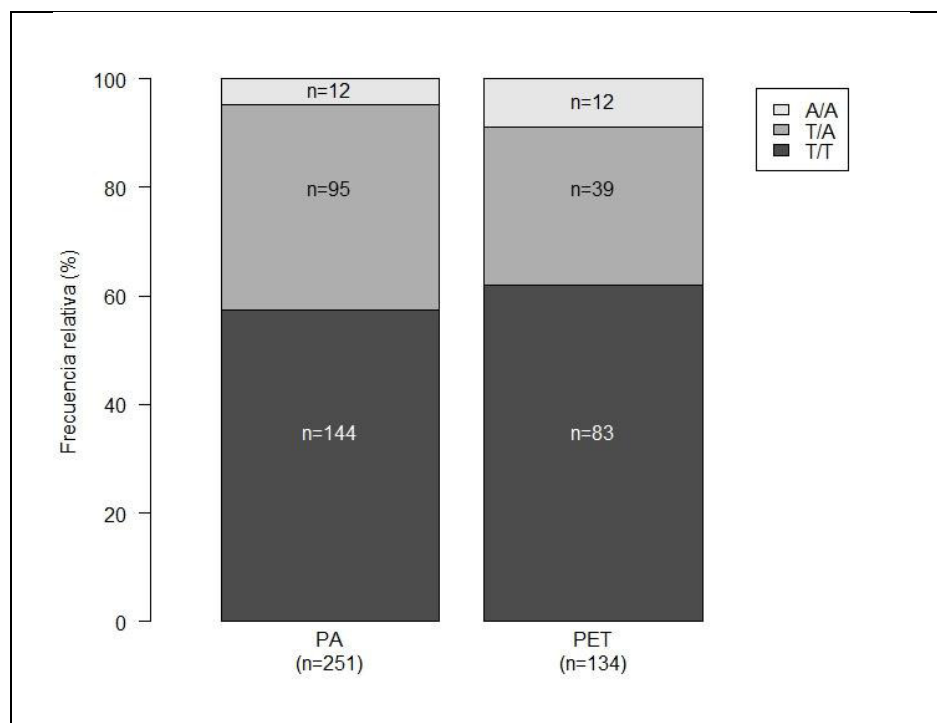


## 10.8 ANEXO 8: Frecuencias alélicas y genotípicas del SNP rs4073

**Tabla 22.** Frecuencias alélicas y genotípicas del SNP rs4073

Genotipos	PA		PET	
	n (%)	Frecuencias alélicas	n (%)	Frecuencias alélicas
		A      T		T      C
		0.24   0.76		0.24   0.76
T/T	144 (57)		83 (62)	
T/A	95 (38)		39 (29)	
A/A	12 (5)		12 (9)	
$\chi^2$	0.539		4.874	
HWE*	0.600		0.031	

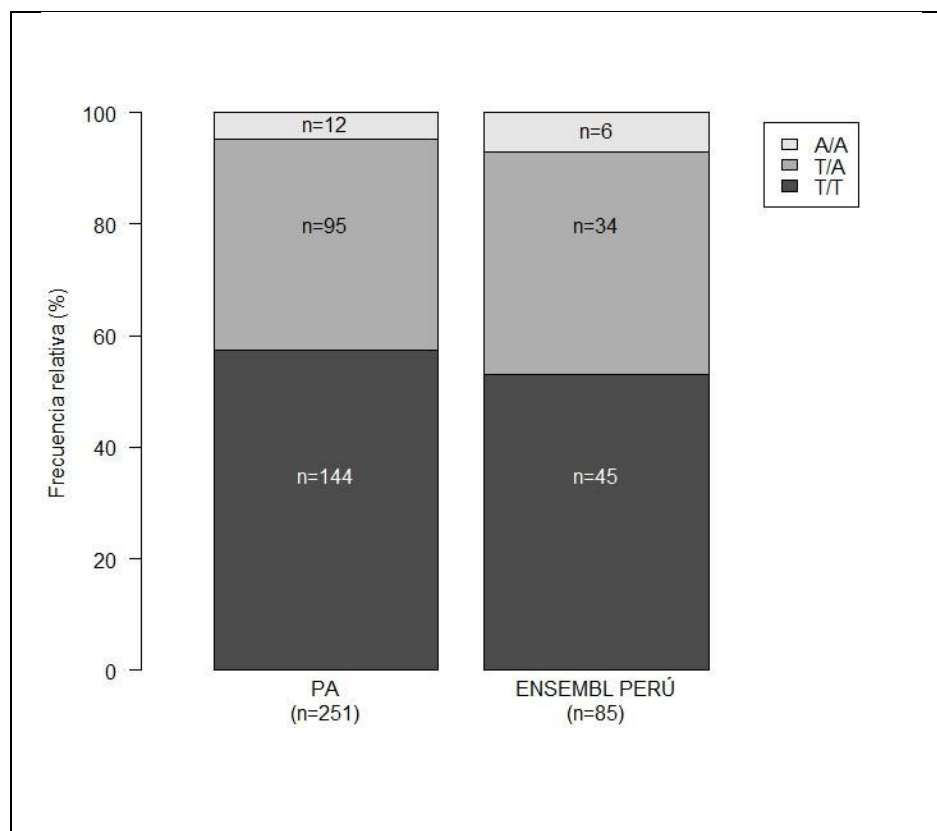
$\chi^2$ : Valor de Chi-cuadrado, EHW: Equilibrio de Hardy-Weinberg.



**Figura 26.** Distribución de las frecuencias genotípicas entre los grupos PA y PET del SNP rs4073.

No hay diferencias estadísticamente significativas (Prueba Chi-cuadrado de Pearson. *p-value*: 0.0968).

10.9 ANEXO 9: Comparación de frecuencias genotípicas del SNP rs4073 con frecuencias reportadas por el proyecto Ensembl

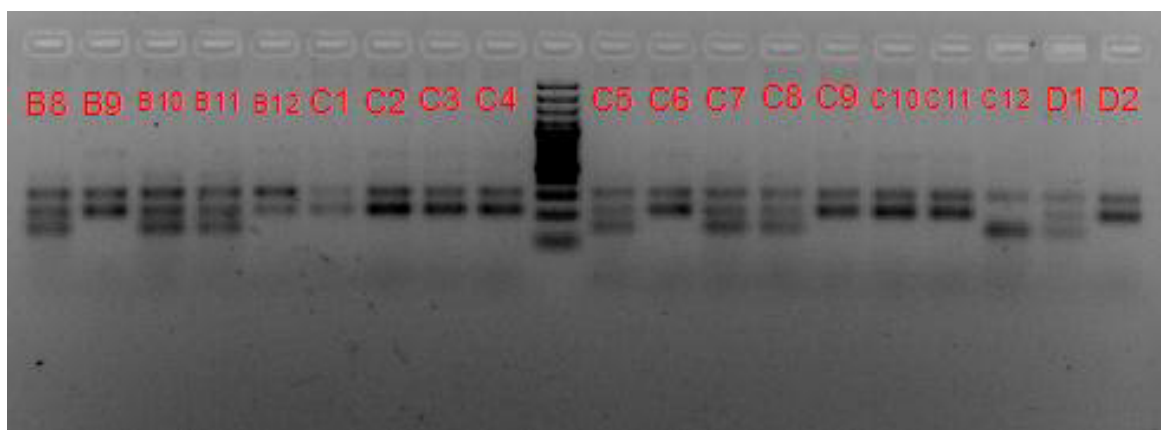


**Figura 27.** Comparación de las frecuencias genotípicas del grupo PA con las del proyecto Ensembl para el SNP rs4073.

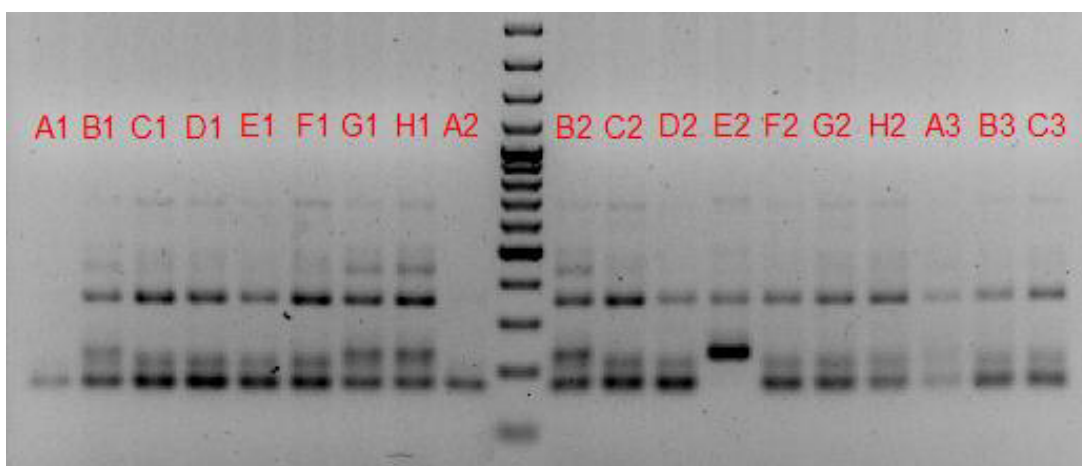
No hay diferencias estadísticamente significativas (Prueba de Fisher, p-value: 0.634).

#### 10.10 ANEXO 10: Genotipificación del SNPs rs2227543 y rs4073

rs2227543



rs4073



**Figura 28.** Genotipificación de los SNPs rs2227543 y rs4073

- Es un trabajo en ciego por lo cual las muestras genotipificadas cuentan con código pero todavía sin la información si pertenecen al grupo PA o PET.